Bayesian Analysis of Multiway Tables in Association Studies: A Model Comparison Approach

Xiaoquan Wen*

Department of Biostatistics, University of Michigan

Abstract

We consider the problem of statistical inference on unknown quantities structured as a multiway table. We show that such multiway tables are naturally formed by arranging regression coefficients in complex systems of linear models for association analysis. In genetics and genomics, the resulting two-way and three-way tables cover many important applications. Within the Bayesian hierarchical model framework, we define the structure of a multiway table through prior specification. Focusing on model comparison and selection, we derive analytic expressions of Bayes factors and their approximations and discuss their theoretical and computational properties. Finally, we demonstrate the strength of our approach using a genomic application of mapping tissue-specific eQTLs (expression quantitative loci).

1 Introduction

In genetics, it is now well known that DNA mutations are widely-spread on genomes (Balding (2006)); their effects on phenotypes may vary largely under different environmental exposures (Hunter (2005)); and a single genetic variant may impact multiple phenotypes through gene networks (Hodgkin (1998)). To comprehensively understand the role of genetic variants in biological processes, modern-day genetic studies have been trending towards designing experiments to interrogate genetic variants genome-wide and investigate their associations with multiple phenotypes under various environmental conditions (Dimas *et al.* (2009)). Ideally, the scientific findings from such experiments should be summarized in *multiway* tables, in which each entry characterizes the magnitude of genetic association (i.e. genetic effect) of a particular variant, under a specific environmental condition, with respect to a unique phenotype measurement. In practice, because these quantities are not directly observed, efficient statistical inference on full multiway tables becomes an important goal for analyzing genetic data.

Motivated by genetic applications, this paper discusses the general problem of statistical inference on unobserved multiway tables in association analysis. Firstly, we show that multiway tables are naturally formulated from a rather general system of linear models by re-arranging relevant regression coefficients. This linear system, including commonly used multiple linear regression and multivariate linear regression models as special cases, is adequate to address a wide range of genetic applications. It is worth noting that our setup differs from most existing work on multiway (tensor) data analysis in a significant

*xwen@umich.edu

way. Most existing approaches (Kolda and Bader (2009), Hoff (2010)) focus on the multi-dimensional structure of *observed* data, as one of the most common goals is to identify their lower-dimensional representations. In our setting, multiway structures are built of *unobserved* regression coefficients and our interest is to infer the multiway tables in their pre-defined dimensions.

In this paper, we emphasize simultaneous inference on multiway tables. In comparison, a naïve approach would fill out multiway tables one entry at a time (e.g. by fitting separate simple linear regressions) while temporarily ignoring the existence of other entries. This method is not only statistically inferior but also possibly defies the purpose of the specific study design. For example, in meta-analysis of genetic association studies, it is only sensible to jointly analyze genetic associations across all participating studies (it is easy to see that the genetic effects of a single variant across multiple studies form a one-dimensional array). Although the limitation of the naïve method is generally well understood, most currently available joint inference approaches only deal with the special case of one-dimensional multiway tables. In statistical genetics literature, many variable selection methods, both in Bayesian (Servin and Stephens (2007), Fridley (2009), Wilson *et al.* (2010), Guan and Stephens (2011)) and Frequentist (Hoggart *et al.* (2008), Wu *et al.* (2009)) domains, have been proposed to identify multiple relevant mutations with respect to a single trait. In meta-analysis and detecting G$\times$E interaction, single variant association testings across multiple subgroups have been thoroughly studied (Lebrec *et al.* (2010), Han and Eskin (2011), Wen and Stephens (2011)) and widely applied. Methodologies for understanding the impacts of genetic variants on multiple phenotypic traits are being developed, but most methods available now propose analyzing one single genetic variant at a time (Stephens (2010)). The generalization of multiway tables not only includes all above applications as special cases, it also enables us to consider a much richer set of important scientific applications. One such example is the fine-mapping of genetic associations in a meta-analysis setting, which aims to identify possibly multiple phenotype-associated mutations by taking into account linkage disequilibrium (LD) and integrating evidence across multiple studies. This problem is most naturally formulated as an inference on a two-way table of genetic effects, with one dimension including multiple candidate variants and the other dimension consisting of multiple participating studies.

In this paper, we take a Bayesian hierarchical model approach to define the complex structure of multiway tables via prior specification and deal with potential correlations in observed data through likelihood functions. This separation makes it conceptually easy to apply the proposed statistical framework in context-dependent settings: we argue that the priors on the multiway effects can be "generic" depending solely on the scientific inquires of interest, whereas only likelihood functions should be adjusted for different experimental designs and sampling procedures.

Our primary focus in inference is to perform model comparison and selection (with hypothesis testings as special cases). In our settings, different candidate multiway tables are characterized by their priors and typically non-nested. In the Bayesian framework, comparing non-nested hierarchical models is straightforward by utilizing Bayes factors. In the following proceeding sections, we derive Bayes factors and their approximations for multiway tables. We discuss both the theoretical and the computational properties of our analytic results and demonstrate their efficiency in applications of Bayesian model selections.

2 Linear Systems for Modeling Multiway Tables

Linear models are effective statistical tools for association analysis. In this section, we describe a general linear model system where regression coefficients naturally form multiway tables.

2.1 System of Simultaneous Multivariate Linear Regressions (SSMR)

A system of simultaneous multivariate linear regressions (SSMR) consists of a set of s multivariate linear regression equations, i.e.,

$$Y_i = X_i B_i + E_i,\ E_i \sim \mathrm{MN}\left(0, I, \Sigma_i\right),\quad i = 1, \ldots, s, \tag{2.1}$$

where "MN" denotes a matrix-variate normal distribution and each linear equation describes one of s non-overlapping subgroups of observed data.

For subgroup i with n_i subjects, Y_i is an $n_i \times r$ matrix with each row representing the r quantitative measurements from one subject; $X_i = (X_{c,i}\ X_{g,i})$ is an $n_i \times (q_i + p)$ design matrix, in which $X_{g,i}$ represents the data matrix of p explanatory variables of interest (e.g. genotypes of interrogated genetic variants) and $X_{c,i}$ represents the data of q_i additional variables (including the intercept) to be controlled for; B_i is a $(q_i + p) \times r$ matrix of regression coefficients and it can be further decomposed into $B_i = \begin{pmatrix} B_{c,i} \\ B_{g,i} \end{pmatrix}$ in which matrices $B_{g,i}$ $(p \times r)$ and $B_{c,i}$ $(q_i \times r)$ contains the regression coefficients for explanatory and controlled variables respectively; finally, E_i is an $n_i \times r$ matrix of residual errors where each row vector is assumed to be independent and identically distributed as $\mathrm{N}(\mathbf{0}, \Sigma_i)$.

Although the same set of r response variables and p explanatory variables are assumed to be measured in all s subgroups, we allow each linear model to control for a different set of covariates. Furthermore, the residual errors are assumed independent across subgroups. In addition, we denote $\boldsymbol{Y} := \{Y_1, \ldots, Y_s\}$, $\boldsymbol{X} := \{X_1, \ldots, X_s\}$ and $\boldsymbol{\Sigma} := \{\Sigma_1, \ldots, \Sigma_s\}$ (throughout the paper, we refer to $\boldsymbol{\Sigma}$ as "residual error variances").

It should be clear that in the SSMR, each $B_{g,i}$ forms a two-dimensional slice and a three-way table of interest for inference can be constructed by joining all s slices. Correspondingly, the resulting three-way table can be "unfolded" into a one-dimensional vector, denoted by $\boldsymbol{\beta}_g$, which is mathematically convenient to work with. More specifically, we define $\boldsymbol{\beta}_g := \begin{pmatrix} \mathrm{vec}(B'_{g,1}) \\ \vdots \\ \mathrm{vec}(B'_{g,s}) \end{pmatrix}$ and similarly, $\boldsymbol{\beta}_c := \begin{pmatrix} \mathrm{vec}(B'_{c,1}) \\ \vdots \\ \mathrm{vec}(B'_{c,s}) \end{pmatrix}$.

To perform Bayesian inference based on the SSMR model, we assign prior distributions for $\boldsymbol{\beta}_g, \boldsymbol{\beta}_c$ and $\boldsymbol{\Sigma}$. The vectorized multiway table $\boldsymbol{\beta}_g$ is of primary interest, for which we assume a multivariate normal prior,

$$\boldsymbol{\beta}_g \sim \mathrm{N}(\mathbf{0}, W_g). \tag{2.2}$$

The Variance-covariance matrix W_g plays a central role in our framework and we defer the detailed discussions to section 3. For the regression coefficients of controlled variables, we assume

$$\boldsymbol{\beta}_c \sim \mathrm{N}\left(\mathbf{0}\ ,\ \Psi_c\right), \tag{2.3}$$

where matrix Ψ_c is assumed diagonal. When performing inference, we consider the limiting condition $\Psi_c^{-1} \to 0$. Furthermore, we assume $\boldsymbol{\beta}_g$ and $\boldsymbol{\beta}_c$ are *a priori* independent. Thus, the vector containing all the regression coefficients, defined by $\boldsymbol{\beta}_{\text{sys}} := \begin{pmatrix} \boldsymbol{\beta}_c \\ \boldsymbol{\beta}_g \end{pmatrix}$, has the following prior distribution

$$\boldsymbol{\beta}_{\text{sys}} \sim \mathrm{N}(\mathbf{0}, \Psi_c \oplus W_g), \tag{2.4}$$

where "$\oplus$" denotes the matrix operation of direct sum. For the residual error variances, we assign an independent inverse Wishart prior for each composing Σ_i, i.e.,

$$\Sigma_i \sim \mathrm{IW}(H_i, m_i). \tag{2.5}$$

In the special case that $r = 1$, the composing multivariate linear regressions degenerate to multiple linear regressions , accordingly, the univariate version of the inverse Wishart distribution reduces to an inverse Gamma distribution.

2.2 Special Cases and Applications in Genetics

The SSMR model is a generalization of a class of linear systems (e.g., when $s = 1$ and $r = 1$, it becomes a multiple linear regression model). Along with its special cases, the SSMR model is capable of handling various applications involving most two-way and some three-way tables in applications of genetics and genomics.

- *Multivariate Linear Regression (MVLR).* When a data set contains only a single subgroup, i.e. $s = 1$, the SSMR model reduces to a single multivariate linear regression (MVLR) model. In case that a single set of unrelated individuals are sampled from the population, the MVLR model is appropriate for at least two major applications in genetics. The first application studies genetic associations of multiple genetic variants with respect to multiple phenotypes; whereas the second application investigates genetic associations of multiple genetic variants with a single phenotype but in different environmental conditions (i.e., a specific study design for investigating gnen-environment interactions). The data application we show in this paper is an example of the latter case.

- *System of Simultaneous Linear Regressions (SSLR).* This special case arises if each composing multivariate linear regression reduces to a multiple linear regression, i.e. $r = 1$. In association studies of multiple genetic variants with a single phenotype, unrelated individual samples can form non-overlapping subgroups in the following scenarios. The first scenario arises in meta-analyses, where each subgroup represents a participating study. In the second possible scenario, gene-environment interaction is of interest and for each environmental condition, a unique subset of individuals is sampled.

The general SSMR model is also uniquely important for many genetics/genomics applications. One such example is the meta-analysis of genetic variants with respect to multiple phenotypes. In this case, the point of interest is typically the complete three-way table of genetic effects (although when genetic variants are analyzed one at a time, i.e. $p = 1$, the three-way table degenerates to a two-way table).

3 Prior Specification for Multiway Tables

In our Bayesian framework, priors play a vital role in both defining the structure and specifying the *a priori* relationships in a multiway table. Prior distributions complement likelihood functions to fully construct hierarchical models which allow information to be efficiently shared across and within different dimensions of a multiway table. In this section, we discuss some of the technical considerations in specifying prior distributions for $\boldsymbol{\beta}_g$, the vectorized multiway table, and illustrate the strength and the modeling flexibility of our Bayesian framework in handling the complex structures of multiway tables.

Based on the Bayesian principle, in considering priors alone, we argue that the scientific problems in hand dictate the specification of priors, whereas other factors like data collection mechanisms should have little or no influence. By this reasoning, we conclude that the prior specifications for different experimental designs should remain the same, as long as they all aim at the same scientific inquiries. For example, as we have shown in section 2.2, for investigating gene-environment interactions, different sampling schemes can lead to either the SSLR or the MVLR model; however, this difference should not vary our *a priori* expectation on the potential underlying gene-environment interactions. Also as a consequence, the conjugacy of the priors, which is typically linked to some specific form of likelihood function, should not be our top consideration at this stage of model construction.

In this work, we limit ourselves to the class of multivariate normal prior distributions represented by (2.2). Thus, a fully specified covariance matrix W_g completes the prior distribution. Although this approach may appear oversimplified, we show that this class of priors can be indeed very flexible and useful in solving a wide range of interesting scientific problems in genetics and genomics.

It is worth pointing out that the only technical requirement for W_g is its positive semi-definiteness. In particular, we emphasize that prior covariance matrices can be rank-deficient. The singularity of W_g matrix directly reflects some linear restrictions on elements in a multiway table and it is extremely useful to describe some candidate models residing in a lower-dimensional space. For example, in a fixed-effect meta-analysis, effects of a common variable are assumed to be the same across different studies. In our framework, this linear restriction can be represented by a singular W_g matrix (appendix A.3). In another example, to assert a particular entry of a multiway table is exactly 0, we simply zero-out the corresponding row and column in the W_g matrix, which again leads to the resulting W_g being singular. It is easy to see, from a generative model point of view, the linear restrictions imposed by such singular normal prior distributions are enforced in the posterior distributions through Bayesian computation.

3.1 Alternative Parametrization of W_g

To start this section, we first formally define the *skeleton* of a multiway table:

DEFINITION 1. *The skeleton of a multiway table is a binary-valued multi-dimensional array whose*

layout and structure are identical to the original multiway table. Each entry in the skeleton indicates if the corresponding element in the original multiway table is non-zero.

It should be noted in many model/variable selection and hypothesis testing problems, the skeleton of a multiway table, rather than the full table, is of primary interest. In these cases, it is helpful to re-parametrize a W_g matrix by

$$W_g = (\Gamma_g, \Lambda_g), \tag{3.1}$$

where Γ_g is a binary matrix with the same row and column layout as matrix W_g and its entries of value 1 mapping the non-zero entries of W_g; and $\Lambda_g = \{w_{ij}\}$ represents the indexed set of actual values of corresponding non-zero entries in W_g. Because of the one-to-one mapping between the entries of a multiway table and the diagonal elements of the corresponding Γ_g matrix, the information of the skeleton is fully conveyed in the Γ_g matrix. In addition, the Γ_g matrix also carries information on *a priori* correlation structure in a multiway table, which is typically also functionally determined by its skeleton in a context-dependent manner (we show by examples in sections 3.2, 5 and appendix A). Thus, there always exists a bijection (whose exact form is context-dependent) between a skeleton and its corresponding Γ_g matrix.

Furthermore, this formulation yields a principled way to specify a hyper-prior distribution on matrix W_g , i.e.,

$$p(W_g) = p(\Lambda_g|\Gamma_g)\ \Pr(\Gamma_g). \tag{3.2}$$

3.2 Example: Prior Specification in Genetic Applications

This section demonstrates the construction of W_g in genetic and genomic applications. In particular, the emphasis is on the specifications of matrix Γ_g and the conditional distribution $\Lambda_g \,|\, \Gamma_g$ for a given skeleton in such context.

In genetic applications of multiway tables, the dimensions that are commonly of interest include

1. The dimension of genetic variants
2. The dimension of phenotypes
3. The dimension of subgroups

The first two dimensions are straightforward to understand, the subgroup dimension is a generic term for which interpretations may vary in different applications: in detecting gene-environment interactions, subgroups correspond to different environmental conditions; whereas in meta-analyses, samples from different studies form subgroups. In genetic applications where all three dimensions are considered, we show that the corresponding W_g can be reasonably decomposed into a block diagonal matrix. In brief, the decomposition is based on the following prior assumptions:

1. the genetic effects of different genetic variants are assumed to be independent *a priori*.

2. for a single genetic variant, its effects on different phenotypes are assumed to be independent *a priori* conditional on the very genetic variant has some effects on those phenotypes.

3. If a variant-phenotype pair is assumed associated, the effects of genetic associations in different subgroups can be modeled by the Bayesian prior proposed by Wen and Stephens (2011) to take into account potential heterogeneity.

It should be noticed that all above prior assumptions are originated from previous works on various one-dimensional array of genetic effects and they are naturally assembled together. In appendix A, we give detailed discussions on these assumptions. Furthermore, we note that this example merely represents a general starting point for relevant genetic applications: with more specific information available, these additional knowledge can be naturally incorporated into our prior constructions in a similar way demonstrated by Stingo *et al.* (2011) and Veyrieras *et al.* (2008).

3.3 Scale-invariant Prior Formulation

In practice, it is often desired that inference results are invariant to scale transformations of response variables. In Bayesian analysis of multiple linear regressions, this property can be achieved by scaling regression coefficients by the residual standard errors and performing inference on the transformed regression coefficients (Servin and Stephens (2007), Wen and Stephens (2011)). Most importantly, the prior distribution is assigned to transformed regression coefficients. Here, we extend this recipe to the SSMR models.

Specifically, we scale each element in $\boldsymbol{\beta}_g$ by its corresponding residual standard error (in the MVLR, the residual standard error for a given regression coefficient is referred to the square root of the corresponding diagonal element in the Σ matrix). More formally, we define a vector of scale-free standardized effects by

$$\boldsymbol{b}_g := S^{-\frac{1}{2}}\boldsymbol{\beta}_g, \tag{3.3}$$

where S is a diagonal matrix permuted from $\oplus_{i=1}^{s}(I \otimes \mathrm{diag}(\Sigma_i))$ to match the order of elements in $\boldsymbol{\beta}_g$. It is easy to see that the resulting $\boldsymbol{b}_g$ is unitless.

Under this setting, a multivariate normal prior distribution $\boldsymbol{b}_g \sim \mathrm{N}(\mathbf{0}, U_g)$ induces a normal prior distribution on $\boldsymbol{\beta}_g$ with mean 0 and

$$W_g = S^{\frac{1}{2}}\, U_g\, S^{\frac{1}{2}}. \tag{3.4}$$

With (3.4), we are able to handle the desired scale-invariant prior formulation as a special case of original scale formulation.

4 Bayes Factors for Multiway Tables

In this section, we derive Bayes factors to facilitate model comparisons and selections for multiway tables in the Bayesian framework. At the most fundamental level, Bayes factors enable us to compare

the supporting evidence from observed data for a set of competitive models (which are not necessarily nested). In case that posterior model probabilities are of direct interest, Bayes factors can be typically utilized as an intermediate computational device for marginal likelihood. For example, in the Metropolis-Hastings algorithm, the Hastings ratio can be evaluated by directly plugging in the Bayes factor of a newly proposed versus current models. There, the computational gain is obtained due to the fact that Bayes factors pre-process the integrals of most nuisance parameters presented in both (proposed and current) models.

In what follows, we discuss the Bayes factors derived from the most general case of the SSMR model assuming the multivariate normal prior (2.2) for a multiway table.

We first give the formal definition of a Bayes factor for a multiway table. Let H_0 denote the trivial null model where $\boldsymbol{\beta}_g \equiv 0$, then for a multiway table characterized by its prior variance W_g, we define a null-based Bayes factor (Liang *et al.* (2008)) as follows

DEFINITION 2. *Under the SSMR model, for a positive definite W_g, the Bayes factor is defined as*

$$\mathrm{BF}(W_g) = \lim_{\Psi_c^{-1} \to 0} \frac{P(\boldsymbol{Y}|\boldsymbol{X}, W_g)}{P(\boldsymbol{Y}|\boldsymbol{X}, H_0)}. \tag{4.1}$$

For technical reasons, the above definition requires W_g to be full-rank, we will extend this definition to allow for singular W_g matrix later in section 4.1.3.

4.1 Analytical Results on Bayes Factors and Their Analytic Approximations

In this section, we discuss our main analytic results on Bayes factors for multiway tables based on the SSMR model. We start by introducing some necessary additional notations: We use $\hat{\boldsymbol{\beta}}_g$ to denote the least square estimate (also the MLE) of $\boldsymbol{\beta}_g$, the vectorized multiway table of interest, and denote its variance by $V_g := \mathrm{Var}(\hat{\boldsymbol{\beta}}_g)$. It should be noted that under the SSMR model, both $\hat{\boldsymbol{\beta}}_g$ and V_g have closed-form expressions: $\hat{\boldsymbol{\beta}}_g$ depends only on observed data $\boldsymbol{X}$ and $\boldsymbol{Y}$, while V_g depends on $\boldsymbol{X}$ and $\boldsymbol{\Sigma}$ (their explicit forms are given in appendix B).

4.1.1 Exact Bayes Factor with Known Residual Error Variances

In the general case of the SSMR model, when the residual error variances are considered known, rather than being assigned priors, the exact Bayes factor for a multiway table defined by a positive-definite W_g can be analytically expressed. We summarize this result in the following lemma:

LEMMA 1. *In the SSMR model, if $\boldsymbol{\Sigma}$ is known, the Bayes factor in definition 2 can be analytically expressed by*

$$\mathrm{BF}(W_g) = |I + V_g^{-1} W_g|^{-\frac{1}{2}} \cdot \exp\left(\frac{1}{2} \hat{\boldsymbol{\beta}}_g' V_g^{-1} \left[W_g (I + V_g^{-1} W_g)^{-1} \right] V_g^{-1} \hat{\boldsymbol{\beta}}_g \right). \tag{4.2}$$

The derivation of lemma 1 is mostly straightforward and we leave the details in appendix B.1.

NOTE 1. *Under the setting of lemma 1 and in addition, provided V_g is also full-rank, the posterior distribution of $\boldsymbol{\beta}_g$ is given by*

$$\boldsymbol{\beta}_g \,|\, \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\Sigma} \sim \mathrm{N}\left((V_g^{-1} + W_g^{-1})^{-1} V_g^{-1} \hat{\boldsymbol{\beta}}_g \,,\, (V_g^{-1} + W_g^{-1})^{-1}\right). \tag{4.3}$$

The quadratic form inside the exponential function of $\mathrm{BF}(W_g)$ *is equivalent to a multivariate Wald statistic computed using the mean and variance from the posterior distribution.*

NOTE 2. *The Bayes factor naturally deals with potential collinearity in predictors. In particular, the evaluation of the Bayes factor does not require the involving design matrices to be full-rank (the details are explained in appendix C). As a result, when highly correlated explanatory variables are included in the model, the Bayes factor can still be stably computed.*

Note 2 is extremely important for genetic applications where genotypes of many spatially closed genetic variants are often highly correlated. In these situations, the desired Bayes factors can be computed straightforwardly without special computational treatments (e.g. Moore–Penrose pseudoinverse).

4.1.2 Approximate Bayes factor with Unknown Residual Error Variances

In more realistic settings, residual error variances are typically unknown and additional integrations with respect to $\boldsymbol{\Sigma}$ are necessary for evaluating desired Bayes factors. Except for a few notable special cases (e.g. the MVLR model with conjugate priors), the exact Bayes factor generally is analytically intractable.

Alternatively, we apply Laplace's method to pursue analytic approximations of exact Bayes factors. Laplace's method has been widely applied in efficient computation of Bayes factors in less complicated models (Kass and Raftery (1995), Raftery (1996), DiCiccio *et al.* (1997), Saville and Herring (2009), Wen and Stephens (2011)). In the case of SSMR model, we identify two different applications of Laplace's method, from both of which the resulting approximate Bayes factors (ABFs) maintain the exact functional form of (4.2), however replace $\boldsymbol{\Sigma}$ with different point estimates.

The first application of the Laplace's method leads to the use of the following point estimate for each $\Sigma_i \in \boldsymbol{\Sigma}$,

$$\hat{\Sigma}_i = \frac{1}{n_i}\left(H_i + (Y_i - X_i\hat{B}_i)'(Y_i - X_i\hat{B}_i)\right), \tag{4.4}$$

and we further denote $\hat{\boldsymbol{\Sigma}} := \{\hat{\Sigma}_1, \ldots, \hat{\Sigma}_s\}$. In (4.4), $\hat{B}_i$ denotes the least square estimate of B_i; H_i is a hyper-parameter from the prior distribution of Σ_i (recall $\Sigma_i \sim \mathrm{IW}(H_i, m_i)$). Note, in the limiting situation such that $H_i \to 0$, $\hat{\Sigma}_i$ becomes the MLE of Σ_i, and it is also asymptotically equivalent to the REML estimate under this setting.

The relevant quantities in (4.2) that are functionally related to $\boldsymbol{\Sigma}$ are V_g and potentially W_g (e.g. in the scale-invariant prior formulation). We denote $\hat{V}_g$ and $\hat{W}_g$ as the corresponding plug-in estimates of V_g and W_g by $\hat{\boldsymbol{\Sigma}}$.

We formally summarize the result of the first Laplace's method in the following proposition:

PROPOSITION 1. *Under the SSMR model, when* $\mathbf{\Sigma}$ *is unknown, applying Laplace's method leads to the following analytic approximation of the Bayes factor*

$$\mathrm{ABF}(W_g) := |I + \hat{V}_g^{-1}\hat{W}_g|^{-\frac{1}{2}} \cdot \exp\left(\frac{1}{2}\hat{\boldsymbol{\beta}}_g'\hat{V}_g^{-1}\left[\hat{W}_g(I + \hat{V}_g^{-1}\hat{W}_g)^{-1}\right]\hat{V}_g^{-1}\hat{\boldsymbol{\beta}}_g\right), \tag{4.5}$$

and

$$\mathrm{BF}(W_g) = \mathrm{ABF}(W_g) \cdot \prod_{i=1}^{s}\left(1 + O(n_i^{-1})\right).$$

Furthermore, provided $n_i \gg p$ *and* $n_i \gg r$ *for* $i = 1, \ldots, s$*, the approximate Bayes factor converges almost surely to the true value, as sample size* $n_i \to \infty$*,* $\forall i$*.*

Proof. See appendix B.2.1 and proof in B.2.2. □

NOTE 3. *Under the conditions considered in the proposition 1,* Σ_i*'s can be accurately estimated by their MLE or RMLE. For the estimator considered in (4.4), it follows that* $\hat{\Sigma}_i \overset{a.s.}{\to} \Sigma_i$ $\forall i$*. This fact, along with lemma 1, provides intuitive explanation for the proposition.*

An alternative application of Laplace's method results in a different analytic approximation of the Bayes factor: each unknown $\Sigma_i \in \mathbf{\Sigma}$ is replaced by the following estimate from the null model,

$$\tilde{\Sigma}_i = \frac{1}{n_i}\left(H_i + (Y_i - X_{c,i}\tilde{B}_{c,i})'(Y_i - X_{c,i}\tilde{B}_{c,i})\right), \tag{4.6}$$

accordingly, we denote $\tilde{\mathbf{\Sigma}} := \{\tilde{\Sigma}_1, \ldots, \tilde{\Sigma}_s\}$. More specifically, $\tilde{B}_{c,i}$ is the least square estimate by fitting the trivial null model, i.e. restricting $B_{g,i} = 0$. By plugging $\tilde{\mathbf{\Sigma}}$, we obtain corresponding estimates of $\tilde{V}_g$ and $\tilde{W}_g$.

The result of the second approximation is formally summarized in below:

PROPOSITION 2. *In the SSMR model, when* $\mathbf{\Sigma}$ *is unknown, an alternative application of Laplace's method leads to the following first-order analytic approximation of the Bayes factor*

$$\mathrm{ABF}^\star(W_g) := |I + \tilde{V}_g^{-1}\tilde{W}_g|^{-\frac{1}{2}} \cdot \exp\left(\frac{1}{2}\hat{\boldsymbol{\beta}}_g'\tilde{V}_g^{-1}\left[\tilde{W}_g(I + \tilde{V}_g^{-1}\tilde{W}_g)^{-1}\right]\tilde{V}_g^{-1}\hat{\boldsymbol{\beta}}_g\right), \tag{4.7}$$

and

$$\mathrm{BF}(W_g) = \mathrm{ABF}^\star(W_g) \cdot \prod_{i=1}^{s}\left(1 + O(n_i^{-1})\right).$$

Proof. See derivation in appendix B.2.3. □

NOTE 4. *Because* $\hat{\boldsymbol{\beta}}_g$ *can also be represented as an analytic function of* $\tilde{B}_c$*, the MLE under the null model, computing* $\mathrm{ABF}^\star(W_g)$ *essentially only requires fitting the trivial null model , which is computationally attractive (details are given in appendix B.2.3).*

Although the two different analytic approximations have the same order of the asymptotic error bounds, in finite sample situations, we expect the good accuracy from $\mathrm{ABF}^\star(W_g)$ only if the true model deviates slightly from the null model. This scenario, in most genetic applications, is typically expected, as most known genetic variants have only very small effects with respect to most phenotypes.

4.1.3 Singular Prior Distribution

So far, our results on Bayes factors and their approximations all require W_g to be full-rank. To extend the definition of Bayes factor for singular W_g, we first define

$$W_g^{\dagger}(\lambda) = W_g + \lambda I, \quad \lambda > 0, \tag{4.8}$$

where W_g is only required positive semi-definite. By this definition, $W_g^{\dagger}(\lambda)$ is guaranteed full-rank. With (4.8), we are now able to extend our definition of Bayes factor for the SSMR model to include singular W_g matrix:

DEFINITION 3. *Under the SSMR model, for a positive semi-definite W_g, the Bayes factor is defined as*

$$\mathrm{BF}(W_g) = \lim_{\lambda \to 0} \mathrm{BF}\left(W_g^{\dagger}(\lambda)\right). \tag{4.9}$$

The definition is based on the following important intuition: Bayes factors are expected to vary very smoothly over a continuum of models. This is not only desired but also critically important for selecting model consistently using Bayes factors.

We have the following result regarding to the existence of the limit:

PROPOSITION 3. *For the SSMR model, the limiting Bayes factors in definition 3 always exist and well-defined, provided that W_g is positive semi-definite.*

In case that $\boldsymbol{\Sigma}$ is known, the proof of proposition 2 is trivial, by noting the the result from lemma 1 does not involve direct inverse of W_g matrix and the matrix sum $(I + V_g^{-1} W_g)$ is always full-rank. If $\boldsymbol{\Sigma}$ is unknown, the arguments are a little involved and we give the full proof for this case in appendix D.

In case of unknown $\boldsymbol{\Sigma}$, Laplace's method can still be applied to derive analytic approximations of the Bayes factor. Briefly, in case of singular W_g matrix, the result for ABF* is unchanged; whereas for ABF, we adjust Laplace's method to take into account the linear restrictions imposed by W_g matrix. This strategy yields a more accurate approximation by using a better estimate of $\hat{\boldsymbol{\Sigma}}$. Nonetheless, the functional form of ABF remains intact. We describe the relevant technical details in appendix D.

With these results, we have extended lemma 1 and propositions 1 and 2 to allow for singular W_g matrix.

4.2 Connections to Frequentist Test Statistics and BIC

4.2.1 Connection to Multivariate Test Statistics

Under the following prior assumptions:

1. $W_g = cV_g$, where c is a positive scalar constant
2. V_g is full-rank

3. $\Sigma_i \sim \text{IW}(H_i, m_i)$ under the limiting conditions, $H_i \to 0$, for $i = 1, \ldots, s$

the resulting prior bears a similarity to Zellner's g-prior (Zellner (1986), Liang *et al.* (2008)). It can be shown that, under this setting,

$$\text{ABF}(W_g) = \left(\sqrt{\frac{1}{1+c}}\right)^{rps} \cdot \exp\left(\frac{1}{2} \cdot \frac{c}{1+c} \cdot T_{\text{wald}}\right), \tag{4.10}$$

and

$$\text{ABF}^\star(W_g) = \left(\sqrt{\frac{1}{1+c}}\right)^{rps} \cdot \exp\left(\frac{1}{2} \cdot \frac{c}{1+c} \cdot T_{\text{score}}\right). \tag{4.11}$$

In particular, T_{wald} and T_{score} represent the multivariate Wald statistic and the Rao's score statistic respectively, both of which are computed based on the SSMR model for testing $H_0 : \boldsymbol{\beta}_g = 0$. Obtaining (4.10) is straightforward and we give the details for (4.11) in appendix E.1.

The important consequence of this connection is that, under this specific form of priors, approximate Bayes factors and the corresponding Frequentist test statistics yield the same rankings for a set of candidate models.

Previously, Wakefield (2009), Johnson (2005, 2008) have identified similar connections between (approximate) Bayes factors and the t-statistic in simple linear regression models. Wen and Stephens (2011) extends their results in a meta-analysis setting. The results we presented here are more general and include all previous findings as special cases.

Albeit the connections, we do not advocate the use of these test statistics as model comparison devices. Especially, caution should be taken when interpreting this prior in specific contexts: for example, Wakefield (2009) and Wen and Stephens (2011) have shown some undesired implications of this prior in genetic applications (e.g., $|W_g|$ is inversely proportionally to sample sizes).

4.2.2 Connection to BIC

Under the suitable conditions, we show (in appendix E.2) that the derived Bayes factor and its approximations can be further approximated by Bayesian Information Criterion (BIC, Schwarz (1978)): Let L_1 and L_0 denote the likelihoods computed from the target model (characterized by W_g) and the null model respectively, it follows that

$$\log \text{BF}(W_g) = (\log L_1 - \log L_0) - \frac{pr}{2} \sum_{i=1}^{s} \log n_i + O(1). \tag{4.12}$$

BIC is asymptotically consistent, meaning that as sample size increases to infinity and under other suitable conditions, BIC selects the fixed true model among a finite set of candidates with probability 1 (Haughton (1988), Schwarz (1978)). Consequently, by (4.12), our Bayes factor and its approximations also enjoy this asymptotic consistency property.

It is worth pointing out that BIC is not a universal approximation of Bayes factors. In our case, BIC (4.12) fails to approximate desired Bayes factors with the advocated error bound in the following two noticeable scenarios:

1. W_g or V_g is singular. Intuitively, in this case, linear constraints on parameter space would change the way that "free" parameters are counted. Nonetheless, it is usually possible to resolve the linear constraints by transformation and re-parametrization.

2. W_g is some function of sample sizes. An interesting example is the prior we studied in the previous section, $W_g = cV_g$, in which $|W_g|$ becomes inversely proportional to sample sizes. For an arbitrary fixed constant c, it is easy to see BIC fails to approximate (4.10) and (4.11). This phenomenon is closely related to the "information paradox" (Liang *et al.* (2008)) and we give detailed explanations in appendix E.2.

It is also clear from the derivation that in high-dimensional settings where the condition $n_i \gg r, p, s$ is not satisfied for some i, then BIC becomes poor approximations to the target (approximate) Bayes factors, i.e. the $O(1)$ error bound is no longer maintained.

4.3 Extension to Non-normal Data

Although we have been working exclusively on quantitative response variables for which normal distributions are assumed, under certain conditions, our results can be extended to non-normal response variables.

Suppose that multiway tables are modeled in a complex system of generalized linear models within the exponential family framework, furthermore, the MLE of the system can be numerically computed for the vectorized effects $\boldsymbol{\beta}_{\text{sys}}$. Following the standard asymptotic maximum likelihood theory, the likelihood of the system can be approximated by a quadratic expansion around its maximum likelihood estimate. This can be equivalently expressed by the following asymptotic approximation,

$$\hat{\boldsymbol{\beta}}_{\text{sys}} \,|\, \boldsymbol{\beta}_{\text{sys}} \sim \text{N}\left(\boldsymbol{\beta}_{\text{sys}}\,,\, \text{Var}(\hat{\boldsymbol{\beta}}_{\text{sys}})\right), \tag{4.13}$$

where $\text{Var}(\hat{\boldsymbol{\beta}}_{\text{sys}})$ is typically approximated using observed Fisher information. Combining with the prior distribution

$$\boldsymbol{\beta}_{\text{sys}} \sim \text{N}(\mathbf{0}\,,\, \Psi_c \oplus W_g), \tag{4.14}$$

it is then straightforward to show that the resulting Bayes factor in this setting maintains the same functional form as (4.2).

For a system of generalized linear models comparable to the SSLR model, (4.13) is straightforward to obtain, similarly to what has been shown by Wen and Stephens (2011). However, challenges remain for finding an appropriate likelihood function to describe correlated non-normal response data in a situation mimicking the MVLR model: for correlated binary data, probit model proposed by Chib and Greenberg (1998) seems a natural fit; however for other general data types, solutions are not generally readily available.

4.4 Bayes factors for Skeletons of Multiway Tables

The Bayes factor of the skeleton of a multiway table, characterized by Γ_g, in the SSMR model can be computed from $\mathrm{BF}(W_g)$ based on the formulation of (3.1), i.e.,

$$\mathrm{BF}(\Gamma_g) = \int p(\Lambda_g \,|\, \Gamma_g)\, \mathrm{BF}(W_g)\, d\,\Lambda_g. \tag{4.15}$$

This formulation provides us the necessary statistical instrument to perform model comparisons in the space of skeletons of multiway tables. In genetic applications, it is often desired to model $p(\Lambda_g \,|\, \Gamma_g)$ by a finite discrete distribution (Stephens and Balding (2009)). In that case, the integration in (4.15) is replaced by a summation.

5 Data Application

In this section, we demonstrate our statistical approaches in mapping tissue-specific eQTLs. eQTLs (expression Quantitative Loci) are genetic variants associated with gene expression levels and they provide biological insights on interplays of genetic variants and gene regulation processes. It also has been widely known that gene regulatory mechanisms may vary significantly in different cell environments. This phenomenon can be indirectly observed by inspecting inconsistent association patterns of relevant eQTLs in different cell types: a putative eQTL may only show associations in some specific cell types when certain regulatory mechanism is active, whereas the very mechanism becomes inactive in some other environments (e.g. prohibited by some regulatory machinery), the association signal is lost.

The data set that we use is originally published by Dimas *et al.* (2009). In their experiment, the investigators genotyped around 500,000 single nucleotide polymorphisms (SNPs) genome-wide in 85 unrelated western European individuals (75 remained after controlling for potential population stratification). Expression levels from this set of individuals were measured genome-wide in primary fibroblasts, Epstein-Barr virus-immortalized B cells (lymphoblastoid cell lines or LCLs), and T cells. The expression data went through quality controls and normalization steps by the original authors, we further select a subset 5,011 genes that are highly expressed in all 3 cell types and perform addional quantile-normalizations for each gene in each cell type. For demonstration purpose, we map eQTLs for each gene separately and narrow the search for eQTLs in the *cis*-region (i.e. the coding region and its close neighborhood) of each gene (note, this is also the strategy adopted in the original publication).

Our goals for mapping tissue-specific eQTLs are two-fold: firstly, we aim to identify eQTLs that show associations in at least one examined cell type (In practice, this is often treated as a hypothesis testing problem); secondly, we are interested in investigating the tissue-specificity of the identified eQTLs. We achieve both goals in a unified model selection/comparison framework by inferring a two-way table of genetic effects for each gene. More specifically, in each two-way table of interest, the columns represent different candidate *cis*-SNPs and the rows represent different cell types. Furthermore, the skeleton of each two-way table is of our primary interest.

Because the expression levels of each gene in different cells are measured from the same set of unrelated individuals, the MVLR model becomes a natural choice for likelihood function. For this data set, the *cis*-region of a gene contains approximately 300 SNPs.

To specify the prior distribution for the two-way table of each gene, we follow the discussion in section 3.2 and assume genetic effects for different SNPs are *a priori* independent. For each SNP, the possible qualitative association pattern with respect to a target gene in 3 cell types (i.e. the skeleton at SNP level) can be represented by a binary 3-vector, $\boldsymbol{\gamma}$. To specify the *a priori* correlation structure, we consider, given that an eQTL is persistently active in a group of cell types, it is likely that the underlying regulatory mechanism is invariant. It is therefore reasonable to assume the genetic effects of the eQTL among this group of cell types are highly correlated. Based on these assumptions, the Γ_g matrix of a two-way table containing p *cis*-SNPs can be precisely represented by its skeletons as

$$\Gamma_g = \oplus_{i=1}^{p} \left(\boldsymbol{\gamma}_i \otimes \boldsymbol{\gamma}_i' \right).$$

Next, we use the scale-invariant prior formulation and apply the Bayesian meta-analysis prior proposed by Wen and Stephens (2011) to specify the conditional distribution $\Lambda_g \,|\, \Gamma_g$. For each SNP, this prior requires two parameters (ϕ, ω) to describe the genetic effects among cell types where it is assumed an active eQTL. In brief, parameter ω characterizes the magnitude of the average genetic effect and parameter ϕ characterizes the heterogeneity of the effects (detailed formulation can be found in appendix A.3). Furthermore, instead of fixing a single set of (ϕ, ω) values for all eQTLs, we assume that (ϕ_i, ω_i) for SNP i, is uniformly sampled from the following set of combinations,

$$L := \{(\phi^{(l)}, \omega^{(l)}) : (0.01, 0.20), (0.01, 0.40), (0.01, 0.80), (0.01, 1.20)\},$$

where the various levels of ω values intend to cover a range of potentially small, modest and large eQTL effects and the relative small ϕ value reflects our prior belief of low heterogeneity across non-zero genetic effects for a given SNP-gene pair due to the hypothesis of invariant regulatory mechanism.

Finally, we specify the prior distribution on possible skeletons (or equivalently Γ_g). By the *a priori* independence assumption on SNPs, for a total of p candidate *cis*-SNPs , we assume the prior of a given skeleton can be expressed as the product of the prior probabilities of its composing columns, i.e.,

$$\Pr(\Gamma_g) = \prod_{i=1}^{p} \Pr(\boldsymbol{\gamma}_i). \tag{5.1}$$

We further treat all SNPs exchangeable and among 2^3 possible configurations of $\boldsymbol{\gamma}_i$, we assign the most probability mass to the null configuration by specifying $\Pr(\boldsymbol{\gamma}_i = (000)) = 0.99$. Our intention here is to encourage overall *sparse* skeletons across *cis*-SNPs, which is motivated by the fact that vast majority of *cis*-SNPs are indeed not associated with the expression levels of a given gene. Given that a SNP is an eQTL, we assume the consistent configuration is mostly likely, i.e. $\Pr(\boldsymbol{\gamma}_i = (111)) = 0.005$, and the rest of the tissue-specific configurations are equally likely by assigning prior probability $0.005/6$ to each of them.

Remark 1. It is important to point out that genome-wide expression-genotype data are typically informative on the distributions of configurations of $\boldsymbol{\gamma}$ and effect size grids in L. In other words, those distributional parameters can be effectively "learned" by pooling information across all genes using some hierarchical model approaches (Veyrieras *et al.* (2008), Maranville *et al.* (2011)). In fact, the hyper-parameters we select here are closely related to the estimations from fitting such a hierarchical model, however these details are not our focus in this paper.

Remark 2. Under this prior specification, the induced marginal prior on the genetic effects of a single SNP can be viewed as a multivariate version of "spike-and-slab" prior (Mitchell and Beauchamp (1988),

George and McCulloch (1993), Ishwaran and Rao (2005)) with the "slab" part also being a mixture of multivariate normals.

Based on the complete model specification, it is straightforward to implement an MCMC algorithm to perform a full posterior inference on Γ_g for each gene. However, this approach is computationally expensive especially when analyzing the data genome-wide. Alternatively, we have designed a greedy algorithm to efficiently search the posterior mode of skeletons. We give the details of this algorithm in appendix F. In brief, this algorithm performs a forward selection by starting at $\Gamma_g = \mathbf{0}$ and proposes to add a new candidate SNP with some non-null configuration at a time in a greedy fashion. The admission of a proposal depends on if the new Γ_g achieves a higher posterior probability. Because of the prior computation is trivial, the essential part of the algorithm involves evaluating Bayes factors for each candidate Γ_g. Since error variances are unknown, we compute $\text{ABF}^\star$ to approximate $\text{BF}(\Gamma_g)$ in our implementation.

We applied this algorithm to Dimas *et al.* (2009) data. Although the sample size is limited and our priors are quite stringent (for identifying eQTLs with small to modest effects), we are able to map 375 eQTLs for the set of 5,011 genes. Among these 375 eQTLs, 291 are called tissue-consistent. We are also able to confidently identify a few genes with multiple *cis*-eQTLs that are not due to linkage disequilibrium (LD), suggesting the involvement of multiple regulatory elements in the transcriptional regulation process of the target genes. Furthermore, most of these multiple eQTLs have different pattern of tissue specificity, verifying the tissue-dependent nature of the gene regulation processes.

We show such an example in Gene YBEY, where three *cis*-eQTLs, rs2075906, rs123298650 and rs2839265, are identified from a total of 236 candidate SNPs. The selection path, the tissue activity configuration of each eQTL and the relevant Bayes factors are shown in Table 1. We note that the genotype correlations among the three called eQTLs are all very weak: the r^2 values for the pairs (rs2075906, rs123298650) and (rs2075906, rs2839265) are close to 0, and the r^2 value between rs123298650 and rs2839265 is merely 0.06. To further examine the called tissue activity configuration, we estimate the marginal effects of each SNP in each individual tissue type and plot the results in Figure 1, which confirms that our inference on cell-type specificity of the eQTLs is quite sensible.

Biologically, we find that both rs2075906 and rs2839265 have been validated as eQTLs by other independent experiments (source: eQTL browser at `http://eqtl.uchicago.edu`). Further bioinformatics analysis (F. Luca, personal communication) reveals that the C allele of SNP rs12329685 disrupts the binding of transcription factor XBP-1, a regulator of YBEY. More interestingly, XBP-1 is only activated by Epstein-Barr virus treatment and therefore presented only in LCL.

6 Conclusion and Discussion

In this paper, we have systematically studied the problem of statistical inference on unobserved multiway tables. we have shown that multiway tables are naturally formed in many commonly used linear model systems and given a generalization in the SSMR model. We have demonstrated the importance of hierarchical modeling and Bayesian prior specification in defining structures of multiway tables. We have also derived analytic expressions of exact and approximate Bayes factors for multiway tables

Step	Candidate	$\log_{10}$ ABF$^{\star}_{\text{single}}$	Configuration	$\log_{10}$ ABF$^{\star}(\Gamma_g)$	Action
1*	rs2075906	4.135	consistent	4.135	Added
2	rs2839156	4.135	consistent	4.191	Not Added
3	rs2839182	3.645	consistent	4.000	Not Added
4	rs2096507	3.591	LCL and T-cell only	5.804	Not Added
5*	rs12329865	3.173	LCL only	7.341	Added
6	rs2839168	3.011	LCL and T-cell only	7.803	Not Added
7	rs914249	2.700	consistent	7.456	Not Added
8	rs1029262	2.698	Fibroblast only	9.583	Not Added
9*	rs2839265	2.693	Fibroblast only	10.485	Added
10	rs2839309	2.611	LCL and T-cell only	10.525	Not Added

Table 1: The selection path generated by the greedy search algorithm running on gene YBEY. ABF$^{\star}_{\text{single}}$ is computed assuming that the candidate SNP is the sole *cis*-eQTL and under its best configuration, while ABF$^{\star}(\Gamma_g)$ assumes a skeleton including the SNPs already selected up to the point along with the candidate SNP in consideration. Some of the candidate SNPs are highly correlated with some existing SNP in the selection set (e.g. rs2839156, rs2839182), thus not added; Others (e.g. rs2096507, rs1029262), although less correlated with the existing set, do not provide strong enough "additional" effects to overcome the prior "penalty".

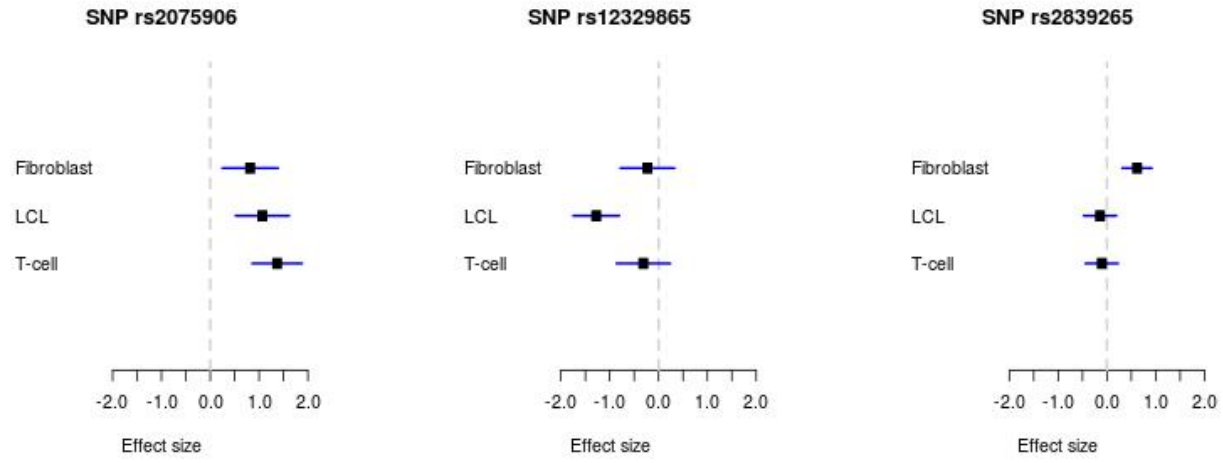

Figure 1: : Examining single SNP effect of each called eQTL in each separate cell type for Gene YBEY. For each SNP in each cell type, we obtain the estimates of the effect size and its standard error by fitting a single linear regression model. The estimated effect sizes and their corresponding 95% confidence intervals are plotted for different cell types in a forest plot for each SNP. SNP rs2839265 has shorter intervals, because its minor allele frequency (0.28) is higher than the frequencies of rs2075906 (0.09) and rs12329865 (0.11)

based on the SSMR model and a general multivariate normal prior assumption. These results provide us a set of statistical instruments to perform efficient Bayesian model comparison and selection for multiway tables. Finally, we have illustrated our statistical approach in mapping multiple potentially

tissue-specific eQTLs.

Albeit the generality of our Bayes factors for arbitrary positive smei-definite W_g matrices, we have placed a great deal of emphasis on context-dependent considerations for W_g and its hyper-prior distributions. In our Bayesian hierarchical models, some important and context-specific properties of multiway tables are articulated through prior specifications. Take sparsity consideration as an example and consider a two-way table in a meta-analysis of multiple variant genetic association studies: it is quite reasonable to assume that only a few genetic variants are associated with the given phenotype; however, conditional on a particular variant is genuinely associated, it is most natural to expect its effects are presented in most studies. Thus, a general (non-structured) sparse assumption on the two-way table becomes inappropriate in this specific settings.

We note in literature several Lasso-type of penalized likelihood methods have been proposed in estimating two-way tables of regression coefficients based on models similar to multivariate linear regressions (Rothman *et al.* (2010), Yin and Li (2011)). We see our approach differs from these methods in three major aspects: firstly and most importantly, as discussed above, our approach requires more context-specific consideration in prior specification than generic sparsity assumption; secondly, our approach separates inferences of skeleton and estimation of regression coefficients, which makes results conceptually easy to interpret; finally, our approach allows correlated predictors in regression models while under the similar scenario Lasso's performance would be affected without adjustments. Nonetheless, we acknowledge that our primary interest in this paper is not *prediction*, for which Lasso-type approaches can be extremely useful.

The general results of approximate Bayes factors are derived through asymptotic approximations. As we have shown, their accuracies ultimately depend on whether RMLE/MLE-type estimators yield reasonably good estimates for unknown residual error variances. When sample sizes are limited and the number of response variables, r, or the number of predictors, p, is modestly large, it is likely that the proposed approximations become inaccurate. This scenario also implies that observed data are not sufficiently informative for $\mathbf{\Sigma}$ and the good inference would inevitably rely on utilization of potential prior information. In the future work, to improve the performance of approximate Bayes factors in small sample situations, it is important to stress both of the following directions: firstly, improve the accuracy of Laplace approximation when likelihood surfaces are flat (some simple numerical strategies proposed in Wen and Stephens (2011) are also applicable here); secondly, construct context-dependent meaningful prior distributions for efficient estimation/integration of $\mathbf{\Sigma}$.

It is worth pointing out that although we have exclusively focused on demonstrations in genetic and genomic applications in this article, our results can be applied in other areas of statistical sciences. The inference and model selection problems in structural equation models (SEM) have been extensively studied (Raftery (1993), Dunson *et al.* (2007)). In particular, a Gaussian directed acyclic graph (DAG) can be represented by a set of linear structural equations. There, the edge set of the graph is typically of interest for inference and is represented by a matrix of regression coefficients. Also linear model systems are also widely used in spatial-temporal modelings, where we also see multiway tables naturally emerge.

A Specification of W_g in Genetic Applications

In this section, we give detailed arguments to sequentially decompose W_g into a block diagonal structure for a three-way table in genetic applications. We emphasize that this particular decomposition should serve as an illustration rather than a general guideline for dealing with genetic applications: especially when additional information becomes available, we can always improve upon it: see Veyrieras *et al.* (2008), Stingo *et al.* (2011) for discussions.

A.1 Prior Decomposition by Genetic Variants

Guan and Stephens (2011) argues that in genetic applications regression coefficients of genetic effects reflect the "causal" effects on the phenotype of interest and there is no obvious reasons to suspect those causal effects among different variants are correlated spatially (note, it is important to distinguish the correlation of observed genotypes and the independence of true genetic effects). Therefore, suppose there are p genetic variants in consideration, we assume their genetic effects are *a priori* independent. Consequently, W_g can be decomposed as

$$W_g = \Phi_1 \oplus \Phi_2 \oplus \cdots \oplus \Phi_p. \tag{A.1}$$

The i-th block matrix Φ_i are specific to the i-th genetic variants, describing its prior genetic effects with different phenotypes in different subgroups. In particular, $\Phi_i = \mathbf{0}$ indicates that the i-th variant is not associated with any phenotype in any subgroup.

A.2 Priors for Multiple Quantitative Traits Associations

Once W_g is decomposed into the level of single genetic variants, further decomposition can be usually achieved in considering their associations with multiple phenotypes.

The interplays between genetic variants and multiple phenotypes are complicated in nature: not only genetic variants can directly affect multiple phenotypes, there also exist interactions between phenotypes through gene networks. Recently, Stephens (2010) proposes a directed acyclic graph (DAG) approach to describe this complex system. Essentially, it first classifies phenotypes into clusters of affected and unaffected by a target genetic variant and decompose Φ_i as follows:

$$\Phi_i = \Theta_{i1} \oplus \cdots \oplus \Theta_{ir}, \tag{A.2}$$

where each block matrix Θ_{ij} characterizes the genetic effects of the i-th variant with respect to phenotype j. If a phenotype j is assumed unaffected, we set $\Theta_{ij} = \mathbf{0}$; otherwise, the non-zero genetic effects are assumed *a priori* independent. Combining with a multivariate linear regression model, such prior specification leads to analyze associations of affected phenotypes with a target genetic variant conditional on unaffected phenotypes.

A.3 Priors for Heterogeneous Effects in Subgroups

Finally, in the dimension of subgroups, the goal is to define heterogeneity of genetic effects of a specific variant with multiple phenotypes in different subgroups.

This problem has been considered by Wen and Stephens (2011) in Bayesian framework. One of the priors they have proposed has the following form to describe the genetic effect of the i-th variant with respect to the j-th phenotype in the k-th subgroup:

$$\beta_{g,ijk} \sim \mathrm{N}(\bar{\beta}_{g,[ij]}\ , \phi^2_{ij}), \tag{A.3}$$

where scalar $\bar{\beta}_{g,[ij]}$ denotes the "average" genetic effect of the variant–phenotype pair across all subgroups considered. Furthermore,

$$\bar{\beta}_{g,[ij]} \sim \mathrm{N}(0, \omega^2_{ij}). \tag{A.4}$$

In this formulation, parameter ϕ_{ij} characterizes the heterogeneity, and ω_{ij} describes the magnitude of prior expected average genetic effect. This prior specification consequently implies that if genetic effects are assumed to present in all subgroups, Θ_{ij} has the following structure (by integrating out $\beta_{g,[ij]}$)

$$\Theta_{ij} = \begin{pmatrix} \phi^2_{ij} + \omega^2_{ij} & \cdots & \omega^2_{ij} \\ \vdots & \ddots & \vdots \\ \omega^2_{ij} & \cdots & \phi^2_{ij} + \omega^2_{ij} \end{pmatrix}. \tag{A.5}$$

The values of (ϕ_{ij}, ω_{ij}) depend on the contexts of the applications. In a meta-analysis, heterogeneity of genuine associations might be assumed small, which leads to specifying small values of ϕ^2_{ij}. In the extreme case of a fixed effect model, ϕ_{ij} is assumed to be exactly 0 (and the resulting Θ_{ij} is singular). However, in detecting G×E interactions, heterogeneity of a true signal is expected to be large and large value of ϕ_{ij} should be considered. The detailed discussion on this topic can be found in Wen and Stephens (2011). This framework also allows to specify genetic effects in certain subgroups are exactly 0, which becomes important in detecting qualitative G×E interaction (e.g. mapping tissue specific eQTLs, detailed in application section).

Each block matrix Θ_{ij} also induces an indicator s-vector, $\boldsymbol{\gamma}_{ij}$, denoting the non-zero associations in all the subgroups considered for the i-th variant and the j-th phenotype. Consequently, the resulting Γ_g for the full multiway table can be mathematically represented by

$$\Gamma_g = \oplus_{i=1}^{p} \left(\oplus_{j=1}^{r} \left(\boldsymbol{\gamma}_{ij} \otimes \boldsymbol{\gamma}'_{ij} \right) \right). \tag{A.6}$$

B Bayes factor Derivation

In this section, we show the derivation of Bayes factors based on the SSMR model.

Recall notations introduced in section 2 in the SSMR model: $\boldsymbol{Y} := \{Y_1, \ldots, Y_s\}, \boldsymbol{X} := \{X_1, \ldots, X_s\} = \{(X_{c,1}, X_{g,1}), \ldots, (X_{c,s}, X_{g,s})\}$ and $\boldsymbol{\Sigma} := \{\Sigma_1, \ldots, \Sigma_s\}$. We further denote the complete collection of regression coefficients by $\boldsymbol{B} := \{B_1, \ldots, B_s\}$.

The likelihood function is given by

$$p(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{B}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{r\sum_{i=1}^{s} n_i}{2}} \cdot \prod_{i=1}^{s} |\Sigma_i|^{-\frac{n_i}{2}} \cdot \mathrm{etr}\left(-\frac{1}{2} \sum_{i=1}^{s} \Sigma_i^{-1} (Y_i - X_i B_i)'(Y_i - X_i B_i) \right) \tag{B.1}$$

where function etr$(\cdot)$ denotes exponential of the trace. Given the least square estimate $\hat{B}_i$ for each composing MVLR, it follows that

$$(Y_i - X_i B_i)'(Y_i - X_i B_i) = (Y_i - X_i \hat{B}_i)'(Y_i - X_i \hat{B}_i) + (B_i - \hat{B}_i)'(X_i' X_i)(B_i - \hat{B}_i). \tag{B.2}$$

Note this decomposition holds even if X_i is rank-deficient (however, $\hat{B}_i$ may not be unique, see McCullagh and Nelder (1989), page 82 for discussions). We further denote $\boldsymbol{\beta}_i := \text{vec}(B_i')$ and $\hat{\boldsymbol{\beta}}_i := \text{vec}(\hat{B}_i')$, and use $\boldsymbol{\beta}_{\text{all}}$ and $\hat{\boldsymbol{\beta}}_{\text{all}}$ to denote the sequentially concatenated vectors of $(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_s)$ and $(\hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_s)$ respectively. The likelihood function (B.1) can be re-written as

$$\begin{aligned} p(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{B}, \boldsymbol{\Sigma}) =& (2\pi)^{-\frac{r\sum_{i=1}^s n_i}{2}} \cdot \prod_{i=1}^s |\Sigma_i|^{-\frac{n_i}{2}} \cdot \text{etr}\left(-\frac{1}{2}\sum_{i=1}^s \Sigma_i^{-1}(Y_i - X_i\hat{B}_i)'(Y_i - X_i\hat{B}_i)\right) \\ & \cdot \exp\left(-\frac{1}{2}\left(\boldsymbol{\beta}_{\text{all}} - \hat{\boldsymbol{\beta}}_{\text{all}}\right)' \Phi \left(\boldsymbol{\beta}_{\text{all}} - \hat{\boldsymbol{\beta}}_{\text{all}}\right)\right), \end{aligned} \tag{B.3}$$

where

$$\Phi = \left(X_1' X_1 \otimes \Sigma_s^{-1}\right) \oplus \cdots \oplus \left(X_s' X_s \otimes \Sigma_s^{-1}\right).$$

Also, by the general case of Gauss-Markov theorem, we note that $\text{Var}(\hat{\boldsymbol{\beta}}_{\text{all}}) = \Phi^{-1}$ (In case that Φ is singular, the Moore–Penrose pseudoinverse is applied).

Although $\boldsymbol{\beta}_{\text{sys}}$ and $\boldsymbol{\beta}_{\text{all}}$ generally differ in the order of the elements, they can be reconciled by a permutation operation, i.e.,

$$P\boldsymbol{\beta}_{\text{all}} = \boldsymbol{\beta}_{\text{sys}}, \tag{B.4}$$

where P is a $(rps + r\sum_i^s q_i) \times (rps + r\sum_i^s q_i)$ permutation matrix. Furthermore, we denote

$$\Omega = P\Phi P,$$

and it can be shown that

$$\text{Var}(\hat{\boldsymbol{\beta}}_{\text{sys}}) = \Omega^{-1}. \tag{B.5}$$

As a result,

$$\begin{aligned} p(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\beta}_{\text{sys}}, \boldsymbol{\Sigma}) =& (2\pi)^{-\frac{r\sum_i^s n_i}{2}} \cdot \prod_{i}^s |\Sigma_i|^{-\frac{n_i}{2}} \cdot \text{etr}\left(-\frac{1}{2}\sum_{i}^s \Sigma_i^{-1}(Y_i - X_i\hat{B}_i)'(Y_i - X_i\hat{B}_i)\right) \\ & \cdot \exp\left(-\frac{1}{2}\left(\boldsymbol{\beta}_{\text{sys}} - \hat{\boldsymbol{\beta}}_{\text{sys}}\right)' \Omega \left(\boldsymbol{\beta}_{\text{sys}} - \hat{\boldsymbol{\beta}}_{\text{sys}}\right)\right). \end{aligned} \tag{B.6}$$

B.1 Bayes Factor for Known $\boldsymbol{\Sigma}$

With $\boldsymbol{\Sigma}$ known, the marginal likelihood $p(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\Sigma})$ can be evaluated analytically, i.e.,

$$p(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\Sigma}) = \int p(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\Sigma}, \boldsymbol{\beta}_{\text{sys}}) p(\boldsymbol{\beta}_{\text{sys}})\, d\boldsymbol{\beta}_{\text{sys}}. \tag{B.7}$$

Recall the prior distribution

$$\boldsymbol{\beta}_{\text{sys}} \sim \text{N}(\mathbf{0}, \Psi_c + W_g).$$

Assuming W_g is full-rank, the integration yields

$$\begin{aligned} p(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\Sigma}) =& (2\pi)^{-\frac{r\sum_i^s n_i}{2}} \cdot \prod_i^s |\Sigma_i|^{-\frac{n_i}{2}} \cdot |W_g|^{-\frac{1}{2}} \cdot |\Psi_c|^{-\frac{1}{2}} \cdot |\Omega + \Psi_c^{-1} \oplus W_g^{-1}|^{-\frac{1}{2}} \\ & \cdot \exp\left(-\frac{1}{2} \hat{\boldsymbol{\beta}}'_{\text{sys}} \Omega \left(\Omega^{-1} - (\Omega + \Psi_c^{-1} \oplus W_g^{-1})^{-1} \right) \Omega \hat{\boldsymbol{\beta}}_{\text{sys}} \right) \\ & \cdot \text{etr}\left(-\frac{1}{2} \sum_i^s \Sigma_i^{-1} (Y_i - X_i \hat{B}_i)'(Y_i - X_i \hat{B}_i) \right), \end{aligned} \tag{B.8}$$

To further simplify (B.8), we decompose Ω into the following block matrix

$$\Omega = \begin{pmatrix} \Omega_c & \Omega_f \\ \Omega'_f & \Omega_g \end{pmatrix},$$

where Ω_c and Ω_g match the the dimensions of matrices Ψ_c and W_g respectively. By (B.5), it follows that

$$V_g^{-1} = \Omega_g - \Omega'_f \Omega_c^{-1} \Omega_f. \tag{B.9}$$

Let

$$\mathcal{U} = \Omega_g - \Omega'_f (\Omega_c + \Psi_c^{-1})^{-1} \Omega_f + W_g^{-1},$$

and it follows that

$$|\Omega + \Psi_c^{-1} \oplus W_g^{-1}| = |\Omega_c + \Psi_c^{-1}| \cdot |\mathcal{U}|. \tag{B.10}$$

Furthermore, the result of the matirx product $\Omega \left(\Omega^{-1} - (\Omega + \Psi_c^{-1} \oplus W_g^{-1})^{-1} \right) \Omega$ can be represented by a block matrix $\begin{pmatrix} A & B \\ B' & D \end{pmatrix}$, where

$$\begin{aligned} A &= \Omega_c \left[I - (\Omega_c + \Psi_c^{-1})^{-1} \Omega_c \right] - \left[I - \Omega_c (\Omega_c + \Psi_c^{-1})^{-1} \right] \Omega_f \mathcal{U}^{-1} \Omega'_f \left[I - (\Omega_c + \Psi_c^{-1})^{-1} \Omega_c \right], \\ B &= \left[I - \Omega_c (\Omega_c + \Psi_c^{-1})^{-1} \right] \Omega_f \mathcal{U}^{-1} W_g^{-1} \\ D &= W_g^{-1} - W_g^{-1} \mathcal{U}^{-1} W_g^{-1} = (\mathcal{U} - W_g^{-1}) - (\mathcal{U} - W_g^{-1}) \mathcal{U}^{-1} (\mathcal{U} - W_g^{-1}). \end{aligned}$$

Although the expressions are fairly complicated, when the limit $\Psi_c^{-1} \to 0$ is taken, $A \to 0$ and $B \to 0$.

The exact same calculation can be carried out for the null model. In the end, we obtain the following marginal likelihood

$$\begin{aligned} P(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\Sigma}, H_0) =& (2\pi)^{-\frac{r\sum_i^s n_i}{2}} \cdot \prod_i^s |\Sigma_i|^{-\frac{n_i}{2}} \cdot |\Psi_c|^{-\frac{1}{2}} \cdot |\Omega_c + \Psi_c^{-1}|^{-\frac{1}{2}} \\ & \cdot \exp\left(-\frac{1}{2} \tilde{\boldsymbol{\beta}}'_c \Omega_c \left(\Omega_c^{-1} - (\Omega_c + \Psi_c^{-1})^{-1} \right) \Omega_c \tilde{\boldsymbol{\beta}}_c \right) \\ & \cdot \text{etr}\left(-\frac{1}{2} \sum_i^s \Sigma_i^{-1} (Y - X_{c,i} \tilde{B}_i)'(Y - X_{c,i} \tilde{B}_i) \right), \end{aligned} \tag{B.11}$$

where $\tilde{\boldsymbol{\beta}}_c$ is the $\hat{\boldsymbol{\beta}}_{\text{sys}}$ estimated under the null model (i.e. restricting $\boldsymbol{\beta}_g \equiv 0$), similarly, $\tilde{B}_i$ is the corresponding $\hat{B}_i$ estimated under the null model for the ith MVLR model. Note the relationship of the least square estimates between the full model and the null model:

$$\tilde{B}_i = \hat{B}_{c,i} + (X'_{c,i}X_{c,i})^{-1}X'_{c,i}X_{g,i}\hat{B}_{g,i}, \tag{B.12}$$

and

$$\begin{aligned}&(Y_i - X_{c,i}\tilde{B}_i)'(Y_i - X_{c,i}\tilde{B}_i) - (Y_i - X_i\hat{B}_i)'(Y_i - X_i\hat{B}_i)\\ &= \hat{B}'_{g,i}\left(X'_{g,i}X_{g,i} - X'_{g,i}X_{c,i}(X'_{c,i}X_{c,i})^{-1}X'_{c,i}X_{g,i}\right)\hat{B}_{g,i}\end{aligned} \tag{B.13}$$

It follows that

$$\begin{aligned}&\text{etr}\left(\frac{1}{2}\sum_i^s \Sigma_i^{-1}\left[(Y_i - X_{c,i}\tilde{B}_i)'(Y_i - X_{c,i}\tilde{B}_i) - (Y_i - X_i\hat{B}_i)'(Y_i - X_i\hat{B}_i)\right]\right)\\ &= \exp\left(\frac{1}{2}\hat{\boldsymbol{\beta}}'_g V_g^{-1}\hat{\boldsymbol{\beta}}_g\right).\end{aligned} \tag{B.14}$$

This also gives the explicit expression for V_g^{-1}, i.e.,

$$V_g^{-1} = \oplus_{i=1}^s \left[\left(X'_{g,i}X_{g,i} - X'_{g,i}X_{c,i}(X'_{c,i}X_{c,i})^{-1}X'_{c,i}X_{g,i}\right)\otimes\Sigma_i^{-1}\right]. \tag{B.15}$$

Finally, by taking the limit $\Psi_c^{-1} \to 0$ and noting

$$\lim_{\psi_c^{-1}\to 0}\mathcal{U} = V_g^{-1} + W_g^{-1}, \tag{B.16}$$

we obtain the final result

$$\text{BF}(W_g) = |I + V_g^{-1}W_g|^{-\frac{1}{2}}\cdot\exp\left(\frac{1}{2}\hat{\boldsymbol{\beta}}'_g V_g^{-1}\left[W_g(I + V_g^{-1}W_g)^{-1}\right]V_g^{-1}\hat{\boldsymbol{\beta}}_g\right). \tag{B.17}$$

B.2 Approximate Bayes Factors for Unknown $\boldsymbol{\Sigma}$

When $\boldsymbol{\Sigma}$ is unknown, we assign independent inverse Wishart priors, $\text{IW}_s(H_i, m_i)$ to each Σ_i and additional integrations are required to compute the marginal likelihood. More specifically, the goal is to evaluate

$$p(\boldsymbol{Y}|\boldsymbol{X}) = \int p(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\Sigma})\prod_i p(\Sigma_i^{-1})\; d\,\Sigma_1^{-1}\ldots d\,\Sigma_s^{-1}, \tag{B.18}$$

where

$$p(\Sigma_i^{-1}) \propto |\Sigma_i^{-1}|^{\frac{m_i - r - 1}{2}}\text{etr}\left(-\frac{1}{2}H_i\Sigma_i^{-1}\right). \tag{B.19}$$

The desired Bayes factor is therefore computed as

$$\text{BF}(W_g) = \lim_{\Psi_c^{-1}\to 0}\frac{\int p(Y|X,\boldsymbol{\Sigma})\prod_i p(\Sigma_i^{-1})\; d\,\Sigma_1^{-1}\cdots d\,\Sigma_s^{-1}}{\int p(Y|X,\boldsymbol{\Sigma},H_0)\prod_i p(\Sigma_i^{-1})\; d\,\Sigma_1^{-1}\cdots d\,\Sigma_s^{-1}}. \tag{B.20}$$

By plugging in (B.8) and (B.11) and noting the cancellation of $|\Psi_c|^{-\frac{1}{2}}$ terms along with the fact that

$$\Omega^{-1} - (\Omega + \Psi_c^{-1} \oplus W_g^{-1})^{-1}$$

is positive definite, it is easy to see that the remaining integrands, both are functions of Ψ_c^{-1}, are bounded. It is then justified by bounded convergence theorem (BCT) to switch limit and integration operations. As a result, we obtain

$$\mathrm{BF}(W_g) = \frac{\int K_{H_a}\, d\,\Sigma_1^{-1} \cdots d\,\Sigma_s^{-1}}{\int K_{H_0}\, d\,\Sigma_1^{-1} \cdots d\,\Sigma_s^{-1}}, \tag{B.21}$$

where

$$\begin{aligned} K_{H_a} = &|I + V_g^{-1}W_g|^{-\frac{1}{2}} \cdot \exp\left(\frac{1}{2}\hat{\boldsymbol{\beta}}_g' \left[V_g^{-1}W_g(I + V_g^{-1}W_g)^{-1}V_g^{-1} - V_g^{-1}\right]\hat{\boldsymbol{\beta}}_g\right) \\ &\cdot \prod_{i=1}^{s} |\Sigma_i^{-1}|^{\frac{n_i+m_i-q_i-r-1}{2}} \cdot \mathrm{etr}\left(-\frac{1}{2}\sum_{i=1}^{s}\Sigma_i^{-1}\left(H_i + (Y_i - X_i\hat{B}_i)'(Y_i - X_i\hat{B}_i)\right)\right), \end{aligned} \tag{B.22}$$

$$K_{H_0} = \prod_{i=1}^{s} |\Sigma_i^{-1}|^{\frac{n_i+m_i-q_i-r-1}{2}} \cdot \mathrm{etr}\left(-\frac{1}{2}\sum_{i=1}^{s}\Sigma_i^{-1}\left(H_i + (Y_i - X_{c,i}\tilde{B}_i)'(Y_i - X_{c,i}\tilde{B}_i)\right)\right), \tag{B.23}$$

It is also important to note that because of (B.14), (B.22) can be alternatively represented as

$$\begin{aligned} K_{H_a} = &|I + V_g^{-1}W_g|^{-\frac{1}{2}} \cdot \exp\left(\frac{1}{2}\hat{\boldsymbol{\beta}}_g' \left[V_g^{-1}W_g(I + V_g^{-1}W_g)^{-1}V_g^{-1}\right]\hat{\boldsymbol{\beta}}_g\right) \\ &\cdot \prod_{i=1}^{s} |\Sigma_i^{-1}|^{\frac{n_i+m_i-q_i-r-1}{2}} \cdot \mathrm{etr}\left(-\frac{1}{2}\sum_{i=1}^{s}\Sigma_i^{-1}\left(H_i + (Y_i - X_{c,i}\tilde{B}_i)'(Y_i - X_{c,i}\tilde{B}_i)\right)\right). \end{aligned} \tag{B.24}$$

Similarly, (B.23) can also be equivalently represented by

$$\begin{aligned} K_{H_0} = &\exp(-\frac{1}{2}\hat{\boldsymbol{\beta}}_g' V_g^{-1}\hat{\boldsymbol{\beta}}_g) \cdot \prod_{i=1}^{s} |\Sigma_i^{-1}|^{\frac{n_i+m_i-q_i-r-1}{2}} \\ &\cdot \mathrm{etr}\left(-\frac{1}{2}\sum_{i=1}^{s}\Sigma_i^{-1}\left(H_i + (Y_i - X_i\hat{B}_i)'(Y_i - X_i\hat{B}_i)\right)\right). \end{aligned} \tag{B.25}$$

Because V_g^{-1} and potentially W_g are both functions of $\boldsymbol{\Sigma}$, the analytic integration of K_{H_a} is generally implausible. Here we approximate both K_{H_a} and K_{H_0} by Laplace's method. Note, although the analytic integration of K_{H_0} is straightforward, it is been shown (Wen and Stephens (2011)) that simultaneously applying Laplace's methods to both K_{H_a} and K_{H_0} achieves better numerical accuracy for desired Bayes factor.

Laplace's method approximates an integral with respect to a $d \times d$ symmetric matrix Z (or equivalently the corresponding half-vectorized $(d+1)d/2$ dimensional vector vech(Z)) in the following way,

$$\int_D h(Z)\exp\left(\, g(Z)\,\right)\, d\,Z \approx (2\pi)^{d(d+1)/4}|H_{\hat{Z}}|^{-1/2}h(\hat{Z})\exp\left(\, g(\hat{Z})\,\right), \tag{B.26}$$

where

$$\hat{Z} = \arg\max_{Z} g(Z),$$

and $|H_{\hat{Z}}|$ is the absolute value of the determinant of the Hessian matrix of the function g evaluated at $\hat{Z}$. The technical requirements on the factorization are that $h(\cdot)$ is smooth and positively valued and $g(\cdot)$ is smooth and obtains its unique maximum in the interior of D. Although different factorization schemes generally achieve different approximation accuracy, the asymptotic error bounds are typically the same (for detailed disscusion, see Butler (2007) chapter 2).

To evaluate the desired Bayes factor, we sequentially apply Laplace's method with respect to each Σ_i^{-1} for both K_{H_a} and K_{H_0}. In what follows, we show the detailed derivations for ABF(W_g) and ABF*(W_g) using two different factorization schemes for K_{H_a}.

B.2.1 Derivation of ABF

The first application of Laplace's method attempts to expand the integrand at the well-known MLE of Σ_i^{-1}. In particular, we sequentially apply this strategy for each Σ_i. We start by factoring K_{H_a} into

$$K_{H_a} = h_a(\Sigma_1^{-1}) \exp\left(g_a(\Sigma_1^{-1}) \right), \tag{B.27}$$

where

$$\begin{aligned} h_a(\Sigma_1^{-1}) = |I + V_g^{-1} W_g|^{-\frac{1}{2}} \cdot \exp\left(\frac{1}{2} \hat{\boldsymbol{\beta}}_g' \left[V_g^{-1} W_g (I + V_g^{-1} W_g)^{-1} V_g^{-1} - V_g^{-1} \right] \hat{\boldsymbol{\beta}}_g \right) \cdot |\Sigma_1^{-1}|^{\frac{m_1 - q_1 - r - 1}{2}} \\ \cdot \prod_{j=2}^{s} |\Sigma_j^{-1}|^{\frac{n_j + m_j - q_j - r - 1}{2}} \cdot \operatorname{etr}\left(-\frac{1}{2} \sum_{j=2}^{s} \Sigma_j^{-1} \left(H_j + (Y_j - X_j \hat{B}_j)'(Y_j - X_j \hat{B}_j) \right) \right) \end{aligned} \tag{B.28}$$

and

$$g_a(\Sigma_1^{-1}) = \frac{n_1}{2} \log |\Sigma_1^{-1}| - \frac{1}{2} \operatorname{trace}\left(\Sigma_1^{-1} \left(H_1 + (Y_1 - X_1 \hat{B}_1)'(Y_1 - X_1 \hat{B}_1) \right) \right). \tag{B.29}$$

This factorization is likely to yield an accurate approximation when $n_1 \to \infty$, because by law of large number

$$\lim_{n_1 \to \infty} \frac{1}{n_1} (Y_1 - X_1 \hat{B}_1)'(Y_1 - X_1 \hat{B}_1) = \Sigma_1. \tag{B.30}$$

This also gives error bound of the approximation $O(\frac{1}{n_1})$. Note that we can factorize

$$|\Sigma_1^{-1}|^{\frac{n_j + m_j - q_j - r - 1}{2}} = |\Sigma_1^{-1}|^{\frac{n_1 + k}{2}} \cdot |\Sigma_1^{-1}|^{\frac{m_j - q_j - r - k - 1}{2}},$$

for some arbitrary k (and here we use $k = 0$). As long as $n \gg k$, the error bound of the approximation does not change.

It is straightforward to show that the unique maximum of $g(\Sigma^{-1})$ is attained at

$$\hat{\Sigma}_1 = \frac{1}{n_1}\left(H_1 + (Y_1 - X_1\hat{B}_1)'(Y_1 - X_1\hat{B}_1)\right). \tag{B.31}$$

Because H_1 is positive definite and $(Y_1 - X_1\hat{B}_1)'(Y_1 - X_1\hat{B}_1)$ is at least positive semi-definite, $\hat{\Sigma}_1$ is invertible. Thus,

$$\widehat{\Sigma_1^{-1}} = \hat{\Sigma}_1^{-1}. \tag{B.32}$$

To compute the Hessian matrix $H_{g_a}(\Sigma_1^{-1})$, we follow (Minka (2000)) and it can be shown that

$$\begin{aligned} H_{g_a}(\Sigma_1^{-1}) =& \frac{\mathrm{d}^2 g_a}{\mathrm{dvech}(\Sigma_1^{-1})\mathrm{dvech}(\Sigma_1^{-1})'} \\ =& -\frac{n_1}{2} D_s' \left(\Sigma_1 \otimes \Sigma_1\right) D_s, \end{aligned} \tag{B.33}$$

where D_s denotes the duplication matrix for $s \times s$ symmetric matrices. When the Hessian is evaluated at $\hat{\Sigma}_1^{-1}$, its absolute determinant results in the following simple form,

$$|H_{g_a}(\hat{\Sigma}_1^{-1})| = 2^{-r} n_1^{r(r+1)/2} |\hat{\Sigma}_1|^{r+1}. \tag{B.34}$$

Similarly, we expand K_{H_0} around $\hat{\Sigma}_1$ based on expression (B.25), which results in the following decomposition,

$$K_{H_0} = h_0(\Sigma_1^{-1}) \exp(\, g_0(\Sigma_1^{-1})\,), \tag{B.35}$$

where

$$\begin{aligned} h_0(\Sigma_1^{-1}) =& \prod_{j=2}^{s} |\Sigma_j^{-1}|^{\frac{n_j + m_j - q_j - r - 1}{2}} \cdot \mathrm{etr}\left(-\frac{1}{2}\sum_{j=2}^{s} \Sigma_j^{-1}\left(H_j + (Y_j - X_j\hat{B}_j)'(Y_j - X_j\hat{B}_j)\right)\right) \\ & \cdot \exp(-\frac{1}{2}\boldsymbol{\beta}_g' V_g^{-1} \boldsymbol{\beta}_g) \cdot |\Sigma_1^{-1}|^{\frac{m_1 - q_1 - r - 1}{2}} \end{aligned} \tag{B.36}$$

and

$$g_0(\Sigma_1^{-1}) = \frac{n_1}{2}\log|\Sigma_1^{-1}| - \frac{1}{2}\mathrm{trace}\left(\Sigma_1^{-1}\left(H_1 + (Y_1 - X_1\hat{B}_1)'(Y_1 - X_1\hat{B}_1)\right)\right). \tag{B.37}$$

Noting that $g_0(\Sigma_1^{-1})$ and $g_1(\Sigma_1^{-1})$ are identical, $\hat{\Sigma}_1$ also uniquely maximizes $g_0(\Sigma_1^{-1})$.

Sequentially, we apply the similar factorization and maximization procedure to all composing Σ_i's and we obtain the following approximation,

$$\mathrm{BF}(W_g) = |I + \hat{V}_g^{-1}\hat{W}_g|^{-\frac{1}{2}} \cdot \exp\left(\frac{1}{2}\hat{\boldsymbol{\beta}}_g' \hat{V}_g^{-1}\left[\hat{W}_g(I + \hat{V}_g^{-1}\hat{W}_g)^{-1}\right]\hat{V}_g^{-1}\hat{\boldsymbol{\beta}}_g\right) \cdot \prod_{i=1}^{s}\left(1 + O(\frac{1}{n_i})\right) \tag{B.38}$$

where $\hat{V}_g^{-1}$ and $\hat{W}_g$ are the corresponding V_g^{-1} and W_g evaluated at $(\hat{\Sigma}_1, \ldots, \hat{\Sigma}_s)$. In particular,

$$\hat{V}_g^{-1} = \oplus_{i=1}^{s}\left[(X_{g,i}'X_{g,i} - X_{g,i}'X_{c,i}(X_{c,i}'X_{c,i})^{-1}X_{c,i}'X_{g,i}) \otimes \hat{\Sigma}_i^{-1}\right]. \tag{B.39}$$

This leads to the final expression of ABF

$$\mathrm{ABF}(W_g) = |I + \hat{V}_g^{-1}\hat{W}_g|^{-\frac{1}{2}} \cdot \exp\left(\frac{1}{2}\hat{\boldsymbol{\beta}}_g' \hat{V}_g^{-1}\left[\hat{W}_g(I + \hat{V}_g^{-1}\hat{W}_g)^{-1}\right]\hat{V}_g^{-1}\hat{\boldsymbol{\beta}}_g\right). \tag{B.40}$$

B.2.2 Proof of the Asymptotic Consistency of ABF

Proof. Under the condition stated in proposition 1, it follows that

$$\hat{\Sigma}_i \xrightarrow{a.s.} \Sigma_i, \text{ for } i = 1, \ldots, s$$

It then follows from the continuous mapping theorem that

$$\text{ABF}(W_g) \xrightarrow{a.s.} \text{BF}(W_g),$$

in which $\text{BF}(W_g)$ is evaluated using the true Σ_i's. □

B.2.3 Derivation of ABF*

The second application of Laplace's method expands K_{H_a} around the MLE of Σ_i^{-1} under the null model. Based on (B.24), we factorize K_{H_a} into,

$$\begin{aligned} h_a(\Sigma_1^{-1}) = &|I + V_g^{-1}W_g|^{-\frac{1}{2}} \cdot \exp\left(\frac{1}{2}\hat{\boldsymbol{\beta}}_g' V_g^{-1} W_g (I + V_g^{-1}W_g)^{-1} V_g^{-1} \hat{\boldsymbol{\beta}}_g\right) \cdot |\Sigma_1^{-1}|^{\frac{m_1 - q_1 - r - 1}{2}} \\ &\cdot \prod_{j=2}^{s} |\Sigma_j^{-1}|^{\frac{n_j + m_j - q_j - r - 1}{2}} \cdot \text{etr}\left(-\frac{1}{2}\sum_{j=2}^{s} \Sigma_j^{-1}\left(H_j + (Y_j - X_{c,j}\tilde{B}_j)'(Y_j - X_{c,j}\tilde{B}_j)\right)\right), \end{aligned} \tag{B.41}$$

and

$$g_a(\Sigma^{-1}) = \frac{n_1}{2}\log|\Sigma_1^{-1}| - \frac{1}{2}\text{trace}\left(\Sigma_1^{-1}\left(H_1 + (Y_1 - X_{c,1}\tilde{B}_1)'(Y_1 - X_{c,1}\tilde{B}_1)\right)\right). \tag{B.42}$$

Next, we factorize K_{H_0} in such a way that $g_a(\Sigma_1^{-1}) = g_0(\Sigma_1^{-1})$, and

$$\begin{aligned} h_0(\Sigma_1^{-1}) = &\prod_{j=2}^{s} |\Sigma_j^{-1}|^{\frac{n_j + m_j - q_j - r - 1}{2}} \cdot \text{etr}\left(-\frac{1}{2}\sum_{j=2}^{s} \Sigma_j^{-1}\left(H_j + (Y_j - X_{c,j}\tilde{B}_j)'(Y_j - X_{c,j}\tilde{B}_j)\right)\right) \\ &\cdot |\Sigma_1^{-1}|^{\frac{m_1 - q_1 - r - 1}{2}} \end{aligned} \tag{B.43}$$

This factorization achieves the desired error bound because of (B.13) and the the following limiting condition

$$\lim_{n_i \to \infty} \frac{1}{n_i}\left(X_{g,i}'X_{g,i} - X_{g,i}'X_{c,i}(X_{c,i}'X_{c,i})^{-1}X_{c,i}'X_{g,i}\right) = Q_i. \tag{B.44}$$

It follows that the unique maximum for g_0 and g_1 is attained at

$$\tilde{\Sigma}_1 = \frac{1}{n_1}\left(H_1 + (Y_1 - X_{c,1}\tilde{B}_1)'(Y_1 - X_{c,1}\tilde{B}_1)\right). \tag{B.45}$$

The remaining calculation is similar to what we have shown in the previous section, and the resulting approximation is given by

$$\text{BF}(W_g) = |I + \tilde{V}_g^{-1}\tilde{W}_g|^{-\frac{1}{2}} \cdot \exp\left(\frac{1}{2}\hat{\boldsymbol{\beta}}_g'\tilde{V}_g^{-1}\left[\tilde{W}_g(I + \tilde{V}_g^{-1}\tilde{W}_g)^{-1}\right]\tilde{V}_g^{-1}\hat{\boldsymbol{\beta}}_g\right) \cdot \prod_{i=1}^{s}\left(1 + O(\frac{1}{n_i})\right), \tag{B.46}$$

where $\tilde{V}_g^{-1}$ and $\tilde{W}_g$ are the corresponding V_g^{-1} and W_g evaluated at $(\tilde{\Sigma}_1, \ldots, \tilde{\Sigma}_s)$. In particular,

$$\tilde{V}_g^{-1} = \oplus_{i=1}^{s} \left[\left(X_{g,i}' X_{g,i} - X_{g,i}' X_{c,i} (X_{c,i}' X_{c,i})^{-1} X_{c,i}' X_{g,i} \right) \otimes \tilde{\Sigma}_i^{-1} \right]. \tag{B.47}$$

Thus, we have obtained an approximation to the desired Bayes factor,

$$\text{ABF}^{\star}(W_g) = |I + \tilde{V}_g^{-1} \tilde{W}_g|^{-\frac{1}{2}} \cdot \exp\left(\frac{1}{2} \hat{\boldsymbol{\beta}}_g' \tilde{V}_g^{-1} \left[\tilde{W}_g (I + \tilde{V}_g^{-1} \tilde{W}_g)^{-1} \right] \tilde{V}_g^{-1} \hat{\boldsymbol{\beta}}_g \right). \tag{B.48}$$

Note the relationship,

$$\text{vec}(\hat{B}_{g,i}') = \left(\left(X_{g,i}' X_{g,i} - X_{g,i}' X_{c,i} (X_{c,i}' X_{c,i})^{-1} X_{c,i}' X_{g,i} \right)^{-1} X_{g,i}' \otimes I \right) \text{vec}[(Y_i - X_{c,i} \tilde{B}_i)'], \tag{B.49}$$

and let

$$\tilde{K}_g = \oplus_{i=1}^{s} \left(X_{g,i} \otimes \tilde{\Sigma}_i^{-1} \right).$$

We then can represent $\text{ABF}^{\star}(W_g)$ using only the estimates from fitting the null model, i.e.,

$$\text{ABF}^{\star}(W_g) = |I + \tilde{V}_g^{-1} \tilde{W}_g|^{-\frac{1}{2}} \cdot \exp\left(\frac{1}{2} \tilde{\boldsymbol{\beta}}_c' \tilde{K}_g \left[\tilde{W}_g (I + \tilde{V}_g^{-1} \tilde{W}_g)^{-1} \right] \tilde{K}_g \tilde{\boldsymbol{\beta}}_c \right), \tag{B.50}$$

C Computational Stability of Bayes Factors

In this section, we show the computational stability of the derived Bayes Factors. The stability issue may arise in practice when highly correlated predictors are included in the models (where V_g might become singular). We show that the derived Bayes Factors and their approximations can still be stably evaluated when the design matrix is or close to rank deficient.

First, we denote

$$G_i = \left(I - X_{c,i} \left(X_{c,i}' X_{c,i} \right)^{-1} X_{c,i}' \right) X_{g,i}, \tag{C.1}$$

and its $p \times n_i$ Moore-Penrose pseudo inverse matrix by G_i^{+}. By the general least square theory, it can be shown (regardless if G_i is full-rank) that

$$\hat{B}_{g,i} = G_i^{+} Y_i, \tag{C.2}$$

$$\hat{\boldsymbol{\beta}}_{g,i} = \text{vec}(\hat{B}_{g,i}') = (G_i^{+} \otimes I) \text{vec}(Y_i') \tag{C.3}$$

$$\tag{C.4}$$

and

$$V_{g,i}^{-1} = \left(G_i' G_i \right) \otimes \Sigma_i^{-1}. \tag{C.5}$$

It is then follows from the general property of Moore-Penrose pseudo inverse, such that

$$\begin{aligned} V_{g,i}^{-1} \hat{\boldsymbol{\beta}}_{g,i} &= \left[\left(G_i' G_i G_i^{+} \right) \otimes \Sigma_i^{-1} \right] \text{vec}(Y_i') \\ &= (G_i' \otimes \Sigma_i^{-1}) \text{vec}(Y_i') \\ &= \text{vec}(\Sigma^{-1} Y_i' G_i). \end{aligned} \tag{C.6}$$

Finally, $V_g^{-1}\hat{\boldsymbol{\beta}}_g$ is computed by sequentially concatenating $V_{g,i}^{-1}\hat{\boldsymbol{\beta}}_{g,i}$ for $i = 1, ..., s$.

Assuming $X'_{c,i}X_{c,i}$ can be inverted in the general sense, we note that in this formulation

1. There is no matrix inversion directly involving $X'_{g,i}X_{g,i}$, which might be rank deficient due to highly correlated predictor variables co-exist in X_g.
2. There is no matrix inversion directly involving W_g.
3. Matrix $(I + V_g^{-1}W_g)$ is guaranteed positive definite.

If $\boldsymbol{\Sigma}$ is unknown and some design matrix X_i is rank deficient, to obtain $\hat{\Sigma}_i$ and evaluate ABF, it becomes inevitable to carry out Moore-Penrose inverse for $X_i'X_i$ in computing $(Y_i - X_i\hat{B}_i)'(Y_i - X_i\hat{B}_i)$. However, in case of ABF*, only $X_{c,i}$ is required for computing $\hat{\Sigma}_i$, which is assumed full-rank.

D Computing Bayes Factors with Singular W_g

We first give the proof for proposition 2 with unknown $\boldsymbol{\Sigma}$.

Proof. By definition (4.9), when W_g is singular and residual error variance Σ is unknown, the desired Bayes Factor is computed by

$$\mathrm{BF}(W_g) = \frac{\lim_{\lambda\to 0}\int K_{H_a}(W_g^\dagger(\lambda))\, d\Sigma_1^{-1} ... \, d\Sigma_s^{-1}}{\int K_{H_0}\, d\Sigma^{-1} ... \, d\Sigma_s^{-1}}, \tag{D.1}$$

where the integrands K_{H_a} and K_{H_0} are defined in (B.22) and (B.23) respectively. It is clear that

$$K_{H_a}(W_g^\dagger(\lambda)) \le \prod_{i=1}^{s}\left[|\Sigma_i^{-1}|^{\frac{n_i+m_i-q_i-r-1}{2}} \cdot \mathrm{etr}\left(-\frac{1}{2}\Sigma_i^{-1}\left(H_i + (Y_i - X_i\hat{B}_i)'(Y_i - X_i\hat{B}_i)\right)\right)\right]. \tag{D.2}$$

Because the RHS is clearly integrable with respect to $\Sigma_1^{-1}, \ldots, \Sigma_s^{-1}$, by bounded convergence theorem (BCT),

$$\mathrm{BF}(W_g) = \frac{\int \lim_{\lambda\to 0} K_{H_a}(W_g^\dagger(\lambda))\, d\Sigma_1^{-1} ... \, d\Sigma_s^{-1}}{\int K_{H_0}\, d\Sigma_1^{-1} ... \, d\Sigma_s^{-1}}. \tag{D.3}$$

It follows that

$$\lim_{\lambda\to 0} K_{H_a}(W_g^\dagger(\lambda)) = K_{H_a}(W_g), \tag{D.4}$$

because K_{H_a} does not involve direct inverse of W_g and matrix sum $(I + V_g^{-1}W_g)$ is guaranteed full rank. Therefore, Bayes factor

$$\mathrm{BF}(W_g) = \frac{\int K_{H_a}(W_g)\, d\Sigma_1^{-1} ... \, d\Sigma_s^{-1}}{\int K_{H_0}\, d\Sigma_1^{-1} ... \, d\Sigma_s^{-1}}. \tag{D.5}$$

exists and well-defined for singular W_g. □

In applying Laplace's method to approximate (D.5), we note that the factorization shown in appendix B.2.1 is still mathematical valid. However, its accuracy is likely worse, because the linear constraints imposed by the singularity of W_g is ignored: under the linear restrictions, the integrand (with respect to $\mathbf{\Sigma}$) is peaked around the constrained estimate of $\mathbf{\Sigma}$ rather than the unconstrained one. Hence, we should modify the factorization based on (B.24) to reflect this fact:

$$\begin{aligned} h_a(\Sigma_1^{-1},\ldots,\Sigma_s^{-1}) =& |I+V_g^{-1}W_g|^{-\frac{1}{2}} \cdot \exp\left(\frac{1}{2}\hat{\boldsymbol{\beta}}_g' V_g^{-1} W_g (I+V_g^{-1}W_g)^{-1} V_g^{-1}\hat{\boldsymbol{\beta}}_g\right) \\ & \cdot \exp\left(-\frac{1}{2}\hat{\boldsymbol{\beta}}_g^{r\prime} V_g^{-1}\hat{\boldsymbol{\beta}}_g^{r}\right) \cdot \prod_{i=1}^{s} |\Sigma_i^{-1}|^{\frac{m_i-q_i-r-1}{2}}, \end{aligned} \tag{D.6}$$

and

$$g_a(\Sigma_1^{-1},\ldots,\Sigma_s^{-1}) = \sum_{i=1}^{s}\left[\frac{n_i}{2}\log|\Sigma_i^{-1}| - \frac{1}{2}\text{trace}\left(\Sigma_i^{-1}\left(H_i + (Y_i - X_i\hat{B}_i^r)'(Y_i - X_i\hat{B}_i^r)\right)\right)\right], \tag{D.7}$$

where $\hat{B}_i^r$ is the least square estimates of B_i subject to the linear constraints and $\hat{\boldsymbol{\beta}}_g^r$ is the corresponding vectorized estimates. The remaining arguments for Laplace's method is the same as we have shown in appendix B.2.1, however $\hat{\Sigma}_i$ is now taking the following form:

$$\hat{\Sigma}_i = \frac{1}{n_i}\left(H_i + (Y_i - X_i\hat{B}_i^r)'(Y_i - X_i\hat{B}_i^r)\right), \tag{D.8}$$

which takes the linear constraints into account.

The derivation for ABF* is invariant under this setting, and the result remains the same.

E Connections with Test Statistics and BIC

E.1 ABF* and Rao's Score Statistic

Following Chen (1983), it is straightforward to derive Rao's score statistic, T_{score} for testing $H_0 : \boldsymbol{\beta}_g = 0$ based on the SSMR model. More specifically,

$$\begin{aligned} T_{\text{score}} &= \sum_{i=1}^{s} \text{vec}[(Y_i - X_{c,i}\tilde{B}_i)']' \left(X_{g,i}'X_{g,i}\otimes\tilde{\Sigma}_i^{-1}\right)\text{vec}[(Y_i - X_{c,i}\tilde{B}_i)'] \\ &= \tilde{\boldsymbol{\beta}}_c'\left[\oplus_{i=1}^{s}\left(X_{g,i}'X_{g,i}\otimes\tilde{\Sigma}_i^{-1}\right)\right]\tilde{\boldsymbol{\beta}}_c. \end{aligned} \tag{E.1}$$

Given that the prior specification $W_g = cV_g$ and W_g is assumed full-rank, it is easy to show that

$$\begin{aligned} & \hat{\boldsymbol{\beta}}_g'\tilde{V}_g^{-1}\left[\tilde{W}_g(I+\tilde{V}_g^{-1}\tilde{W}_g)^{-1}\right]\tilde{V}_g^{-1}\hat{\boldsymbol{\beta}}_g \\ =& \frac{c}{1+c}\cdot\hat{\boldsymbol{\beta}}_g'\tilde{V}_g^{-1}\hat{\boldsymbol{\beta}}_g \\ =& \frac{c}{1+c}\cdot\tilde{\boldsymbol{\beta}}_c'\left[\oplus_{i=1}^{s}\left(X_{g,i}'X_{g,i}\otimes\tilde{\Sigma}_i^{-1}\right)\right]\tilde{\boldsymbol{\beta}}_c. \end{aligned}$$

This yields,

$$\mathrm{ABF}^{\star}(W_g) = \left(\sqrt{\frac{1}{1+c}}\right)^{rps} \cdot \exp\left(\frac{1}{2} \cdot \frac{c}{1+c} \cdot T_{\text{score}}\right)$$

E.2 Connections with BIC

Under the conditions that

1. V_g and W_g are full-rank,
2. $\lim_{n_i \to 0} \frac{\log |W_g|}{n_i} = 0, \forall i$,
3. $n_i \gg p, r, s, \forall i$,

We show that BIC can be derived from the Bayes factor and its approximations based on the SSMR model.

For each consisting multivariate linear regression in MVLR, independent subjects are sampled from a population. Under this setting, it is commonly assumed that

$$\lim_{n_i \to \infty} \frac{1}{n_i} \left(X'_{g,i} X_{g,i} - X'_{g,i} X_{c,i} (X'_{c,i} X_{c,i})^{-1} X'_{c,i} X_{g,i}\right) = Q_i, \tag{E.2}$$

and Q_i is also full-rank. Hence,

$$\lim_{n_i \to \infty} V_g = \oplus_{i=1}^{s} \left[\frac{1}{n_i} \left(Q_i^{-1} \otimes \Sigma_i\right)\right]. \tag{E.3}$$

When $\mathbf{\Sigma}$ is known, as $n_i \to \infty$ for each i, based on (E.3)

$$\lim_{n_i \to \infty, \forall i} \left(I + V_g^{-1} W_g\right) = V_g^{-1} W_g, \tag{E.4}$$

and

$$\lim_{n_i \to \infty, \forall i} \mathrm{BF}(W_g) = |V_g|^{1/2} \cdot |W_g|^{-1/2} \cdot \exp\left(\frac{1}{2} \hat{\boldsymbol{\beta}}'_g V_g^{-1} \hat{\boldsymbol{\beta}}_g\right). \tag{E.5}$$

Note that

$$\lim_{n_i \to \infty} |V_g| = \prod_{i=1}^{s} \left(n_i^{-pr} \cdot |Q_i|^{-r} \cdot |\Sigma_i|^{p}\right), \tag{E.6}$$

and the likelihood ratio

$$L_1 / L_0 = \frac{p(\boldsymbol{Y} | \boldsymbol{X}, \hat{\boldsymbol{B}}, \mathbf{\Sigma})}{p(\boldsymbol{Y} | \boldsymbol{X}, \tilde{\boldsymbol{B}}, \mathbf{\Sigma}, H_0)} = \exp\left(\frac{1}{2} \hat{\boldsymbol{\beta}}'_g V_g^{-1} \hat{\boldsymbol{\beta}}_g\right) \tag{E.7}$$

Finally, we obtain

$$\begin{aligned}\log \mathrm{BF}(W_g) &\approx (\log L_1 - \log L_0) - \frac{pr}{2}\sum_{i=1}^{s}\log n_i + \left(\frac{r}{2}\sum_{i=1}^{s}\log|Q_i| - \frac{p}{2}\sum_{i=1}^{s}\log|\Sigma_i| - \frac{1}{2}\log|W_g|\right) \\ &= (\log L_1 - \log L_0) - \frac{pr}{2}\sum_{i=1}^{s}\log n_i + O(1),\end{aligned} \tag{E.8}$$

which is essentially the BIC.

It is important to note that if V_g or W_g is singular, (E.4) and the rest of the derivation, will not be valid. This is understandable intuitively: the singularity reflects linear constraints on parameter space, and the number of "free" parameters is no longer trivial to count and BIC may not be well defined. Also, if W_g is also a function of sample sizes, then the condition 2 might be violated and the resulting error term may not be of the order $O(1)$. Similarly, condition 3 is also important to ensure the desired error bound for BIC.

When $\mathbf{\Sigma}$ is unknown, in a similar way, it can be shown that

$$\log \mathrm{ABF}(W_g) \approx \frac{1}{2}\hat{\boldsymbol{\beta}}_g'\hat{V}_g^{-1}\hat{\boldsymbol{\beta}}_g - \frac{pr}{2}\sum_{i=1}^{s}\log n_i + \left(\frac{r}{2}\sum_{i=1}^{s}\log|Q_i| - \frac{p}{2}\sum_{i=1}^{s}\log|\hat{\Sigma}_i| - \frac{1}{2}\log|\hat{W}_g|\right), \tag{E.9}$$

and

$$\log \mathrm{ABF}^{\star}(W_g) \approx \frac{1}{2}\hat{\boldsymbol{\beta}}_g'\tilde{V}_g^{-1}\hat{\boldsymbol{\beta}}_g - \frac{pr}{2}\sum_{i=1}^{s}\log n_i + \left(\frac{r}{2}\sum_{i=1}^{s}\log|Q_i| - \frac{p}{2}\sum_{i=1}^{s}\log|\tilde{\Sigma}_i| - \frac{1}{2}\log|\tilde{W}_g|\right). \tag{E.10}$$

Asymptotically,

$$\lim_{n_i\to\infty,\forall i} \hat{\boldsymbol{\beta}}_g'\hat{V}_g^{-1}\hat{\boldsymbol{\beta}}_g \to \hat{\boldsymbol{\beta}}_g V_g^{-1}\hat{\boldsymbol{\beta}}_g. \tag{E.11}$$

Furthermore, it can be shown that

$$\lim_{n_i\to\infty} \tilde{\Sigma}_i = \hat{\Sigma}_i + \hat{B}_{g,i}'Q_i\hat{B}_{g,i}, \tag{E.12}$$

and

$$\lim_{n_i\to\infty,\forall i} \frac{\hat{\boldsymbol{\beta}}_g'\tilde{V}_g^{-1}\hat{\boldsymbol{\beta}}_g}{\hat{\boldsymbol{\beta}}_g'\hat{V}_g^{-1}\hat{\boldsymbol{\beta}}_g} = C, \tag{E.13}$$

where C is a constant. Thus,

$$\hat{\boldsymbol{\beta}}_g'\tilde{V}_g^{-1}\hat{\boldsymbol{\beta}}_g = \hat{\boldsymbol{\beta}}_g'\hat{V}_g^{-1}\hat{\boldsymbol{\beta}}_g + O(1). \tag{E.14}$$

This yields our final results: under the conditions stated

$$\log \mathrm{ABF}(W_g) = (\log L_1 - \log L_0) - \frac{pr}{2}\sum_{i=1}^{s}\log n_i + O(1), \tag{E.15}$$

and

$$\log \mathrm{ABF}^{\star}(W_g) = (\log L_1 - \log L_0) - \frac{pr}{2}\sum_{i=1}^{s}\log n_i + O(1). \tag{E.16}$$

F A Greedy Algorithm for Searching Posterior Mode of Γ_g

In this section, we give the detailed general descriptions of a greedy algorithm for searching the posterior mode of a two-way table.

We first define posterior score

$$S_{\text{post}}(\Gamma_g) := \log_{10} \Pr(\Gamma_g) + \log_{10} \text{BF}(\Gamma_g). \tag{F.1}$$

It should be clear that $\exp(S_{\text{post}}) \propto \Pr(\Gamma_g|Y, G)$.

The algorithm begins by evaluating all Γ_g's containing exactly one SNP and constructing a priority list l:

1. *For SNP $i = 1$ to p:*
   (a) *For subgroup configuration $j = 1$ to s:*
      i. *Construct $\Gamma_{g,ij}$ by a single SNP i with configuration j, compute $S_{\text{post}}(\Gamma_{g,ij})$.*
   (b) *Insert SNP i with the best configuration $j = \arg\max_k S_{\text{post}}(\Gamma_{g,ik})$ to the priority list l to represent SNP i.*
2. *Sort priority list l in a descending order according to the posterior scores of all SNPs.*
3. *Starting with $\Gamma_g = 0$ and $S_{\text{post}}(\Gamma_g) = 0$, go through the list l in the following fashion*
   (a) *Construct a new skeleton, Γ'_g, by adding a new SNP along with its best configuration to current Γ_g, compute $S_{\text{post}}(\Gamma'_g)$.*
   (b) *If $S_{\text{post}}(\Gamma'_g) > S_{\text{post}}(\Gamma_g)$, set $\Gamma_g = \Gamma'_g$, i.e. permanently add the new SNP to the selected set.*
   (c) *Otherwise, leave Γ_g unchanged.*
4. *Report final Γ_g.*

Note that in step 3(b), Γ_g and Γ'_g differs at only one column. Suppose this column has configuration j, by our prior specification on Γ_g,

$$\log_{10} \Pr(\Gamma'_g) - \log_{10} \Pr(\Gamma_g) = \log_{10} \pi + \log(\lambda_j) - \log_{10}(1 - \pi). \tag{F.2}$$

It then becomes equivalent that

$$\begin{aligned} &S_{\text{post}}(\Gamma'_g) > S_{\text{post}}(\Gamma_g) \Leftrightarrow \\ &\log_{10} \text{BF}(\Gamma'_g) - \log_{10} \text{BF}(\Gamma_g) > \log_{10}(1 - \pi) - \log_{10}(\pi) - \log(\lambda_j). \end{aligned} \tag{F.3}$$

This indicates the selection is based on thresholding the Bayes Factors and the threshold is determined by the prior distributions.

References

Balding, D. J. (2006). A tutorial on statistical methods for population association studies. *Nature Review Genetics*, **7**(10), 781–791.

Butler, R. (2007). *Saddlepoint Approximations with Applications*. Cambridge University Press, 1st edition.

Chen, C.-F. (1983). Score Tests for Regression Models. *Journal of the American Statistical Association*, **78**(381), 158–161.

Chib, S. and Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, **85**(2), 347–361.

DiCiccio, T. J., Kass, R. E., Raftery, A., and Wasserman, L. (1997). Computing Bayes Factors by Combining Simulation and Asymptotic Approximations. *Journal of the American Statistical Association*, **92**(439), 903–915.

Dimas, A. S., Deutsch, S., Stranger, B. E., Montgomery, S. B., Borel, C., Attar-Cohen, H., Ingle, C., Beazley, C., Gutierrez Arcelus, M., Sekowska, M., Gagnebin, M., Nisbett, J., Deloukas, P., Dermitzakis, E. T., and Antonarakis, S. E. (2009). Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science*, **325**(5945), 1246–1250.

Dunson, D. B., Palomo, J., and Bollen, K. (2007). Bayesian structural equation modeling. In S.-Y. Lee, editor, *Handbook of Latent Variable and Related Models*, pages 163–188. Elsevier.

Fridley, B. L. (2009). Bayesian variable and model selection methods for genetic association studies. *Genetic Epidemiology*, **33**(1), 27–37.

George, E. I. and McCulloch, R. E. (1993). Variable Selection Via Gibbs Sampling. *Journal of the American Statistical Association*, **88**(423), 881–889.

Guan, Y. and Stephens, M. (2011). Bayesian variable selection regression for genome-wide association studies, and other large-scale problems. *Annals of Applied Statistics*, **5**(3), 1780–1815.

Han, B. and Eskin, E. (2011). Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *The American Journal of Human Genetics*, **8**(3), 586–598.

Haughton, D. M. A. (1988). On the Choice of a Model to Fit Data from an Exponential Family. *The Annals of Statistics*, **16**(1), 342–355.

Hodgkin, J. (1998). Seven types of pleiotropy. *International Journal of Developmental Biology*, **42**, 501–505.

Hoff, P. (2010). Hierarchical multilinear models for multiway data. *arXiv pre-print:1005.5425*.

Hoggart, C. J., Whittaker, J. C., De Iorio, M., and Balding, D. J. (2008). Simultaneous analysis of all snps in genome-wide and re-sequencing association studies. *PLoS Genetics*, **4**(7), e1000130.

Hunter, D. J. (2005). Gene-environment interactions in human diseases. *Nature Review Genetics*, **6**(4), 287–298.

Ishwaran, H. and Rao, J. S. (2005). Spike and Slab Variable Selection: Frequentist and Bayesian Strategies. *Annals of Statistics*, **33(2)**, 730 – 773.

Johnson, V. E. (2005). Bayes factors based on test statistics. *Journal of the Royal Statistical Society - Series B: Statistical Methodology*, **67**(5), 689–701.

Johnson, V. E. (2008). Properties of bayes factors based on test statistics. *Scandinavian Journal of Statistics*, **35**(2), 354–368.

Kass, R. E. and Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, **90**(430), 773–795.

Kolda, T. G. and Bader, B. W. (2009). Tensor decompositions and applications. *SIAM review*, **51**(3).

Lebrec, J. J., Stijnen, T., and van Houwelingen, H. (2010). Dealing with heterogeneity between cohorts in genomewide snp association studies. *Statistical Applications in Genetics and Molecular Biology*, **9**(1), Article 8.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g Priors for Bayesian Variable Selection. *Journal of the American Statistical Association*, **103**(481), 410–423.

Maranville, J. C., Luca, F., Richards, A. L., Wen, X., Witonsky, D. B., Baxter, S., Stephens, M., and Di Rienzo, A. (2011). Interactions between glucocorticoid treatment and cis-regulatory polymorphisms contribute to cellular response phenotypes. *PLoS Genetics*, **7**(7).

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. Chapman and Hall, 2nd edition.

Minka, T. P. (2000). Old and new matrix algebra useful for statistics.

Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian Variable Selection in Linear Regression. *Journal of the American Statistical Association*, **83**(404), 1023–1032.

Raftery, A. E. (1993). Bayesian Model Selection in Structural Equation Models. In K. A. Bollen, editor, *Testing Structural Equation Models*, pages 163–180. Sage Publications.

Raftery, A. E. (1996). Approximate Bayes factors and accounting for model uncertainty in generalised linear models. *Biometrika*, **83**(2), 251–266.

Rothman, A. J., Levina, E., and Zhu, J. (2010). Sparse Multivariate Regression With Covariance Estimation. *Journal of Computational and Graphical Statistics*, **19(4)**, 947–962.

Saville, B. R. and Herring, A. H. (2009). Testing random effects in the linear mixed model using approximate bayes factors. *Biometrics*, **65**(2), 369–376.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6(2)**, 461–464.

Servin, B. and Stephens, M. (2007). Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS genetics*, **3**(7), e114.

Stephens, M. (2010). A unified framework for testing multiple phenotypes for association with genetic variants. *Presented at the 60th Annual Meeting of The American Society of Human Genetics, Washington D.C.*

Stephens, M. and Balding, D. J. (2009). Bayesian statistical methods for genetic association studies. *Nature Review Genetics*, **10**(10), 681–90.

Stingo, F. C., Chen, Y. A., Tadesse, M. G., and Vannucci, M. (2011). Incorporating biological information into linear models: A Bayesian approach to the selection of pathways and genes. *Annals of Applied Statistics*, **5**(3), 1978–2002.

Veyrieras, J., Kudaravalli, S., Kim, S. Y., Dermitzakis, E. T., Gilad, Y., Stephens, M., and Pritchard, J. K. (2008). High-resolution mapping of expression-qtls yields insight into human gene regulation. *PLoS Genetics*, **4**(10).

Wakefield, J. (2009). Bayes factors for genome-wide association studies: comparison with p-values. *Genetic Epidemiology*, **33**(1), 79–86.

Wen, X. and Stephens, M. (2011). Bayesian methods for genetic association analysis with heterogeneous subgroups: from meta-analyses to gene-environment interactions. *arXiv pre-print: 1111.1210*.

Wilson, M. A., Iversen, E. S., Clyde, M. A., Schmidler, S. C., and Schildkraut, J. M. (2010). Bayesian model search and multilevel inference for snp association studies. *Annals of Applied Statistics*, **4**(3), 1342–1364.

Wu, T., Chen, Y., Hastie, T., Sobel, E., and Lange, K. (2009). Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics*, **25**(6), 714–721.

Yin, J. and Li, H. (2011). A sparse conditional Gaussian graphical model for analysis of genetical genomics data. *Annals of Applied Statistics*, **5**(4), 2630 – 2650.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, pages 233–243.